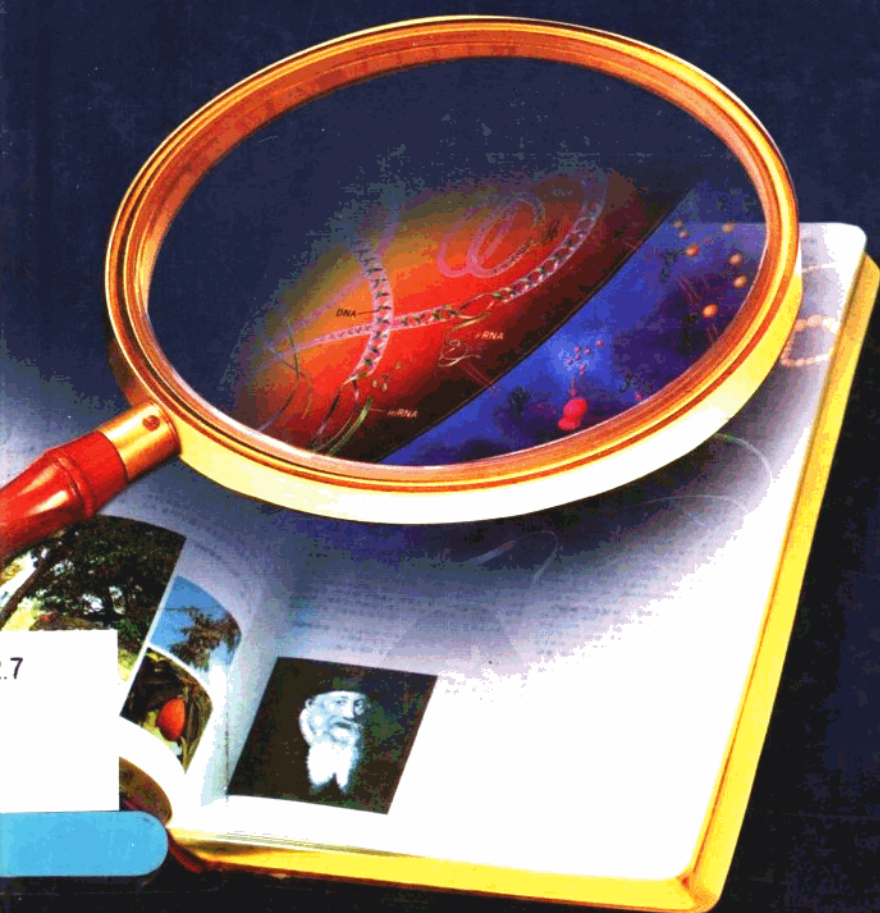


新编科技文献检索

主 编 钱树云

副主编 孙荣林

河海大学出版社



前 言

人类社会已经进入 21 世纪,这是一个科技创新的时代,是一个知识经济的时代,更是一个信息化、网络化的时代。如何在急剧膨胀的知识信息的海洋里,全面、快捷、准确、合理地获取和利用各种对自己有用的信息,已成为新时期大学生以及科技工作人员必须具备的一项技能,也是迫切需要学习和研究的一个问题。本书正是在这样一个背景下,在多年教学实践经验的基础上,结合理工类高校的实际教学情况编写而成的。

全书共分基础篇、化工篇、机电篇和综合篇四个部分。基础篇着重介绍文献手工检索与计算机检索的基础知识;化工篇主要介绍美国《化学文摘》和其他常用的化工类中外文参考与检索工具书、国内外化学文献数据库等内容;机电篇着重介绍美国《工程索引》、英国《科学文摘》、美国《金属文摘》和其他常用的机电类中文参考与检索工具书;综合篇则着重介绍知识产权方面的基础知识与专利文献、标准资料的检索以及国内最权威的 3 种综合数据库等内容。本书最突出的一个特点就是特别注重将各类文献资料的传统手工检索与现代化计算机检索的内容有机地结合在一起。我们相信,通过本书的学习,广大读者一定会获益匪浅。

本书第一章、第七章由钱树云同志编写;第二、第五、第六、第十三章由林玉玲同志编写;第三章的第一、第四节由管春英同志编写;第三章的第二、第三节和第四节及第十二章由刘小华同志编写;金耀同志编写了第八、第九、第十章;孙荣林同志编写了第十一章。

本书的出版得到了江苏石油化工学院教务处史国栋教授和张煌老师的关心与支持。同时,全书在编写过程中也参考了大量同行们的文章、专著,由于篇幅所限,不能一一列出,在此谨表示诚挚的谢意。

因作者水平有限,书中疏漏、错误在所难免,敬请读者批评指正。

编著者

2001 年 7 月

目 录

基 础 篇

第一章 文献检索基础知识.....	1
第一节 概论.....	1
第二节 检索工具和检索语言.....	5
第三节 检索原理、途径、方法和步骤.....	8
第二章 计算机检索基础知识.....	11
第一节 计算机检索概况.....	11
第二节 数据库及检索.....	12

化 工 篇

第三章 《化学文摘》.....	15
第一节 概况.....	15
第二节 内容编排及著录格式.....	16
第三节 CA 的索引	23
第四节 CA 的计算机检索	47
第四章 中文化学化工参考与检索工具书.....	61
第一节 参考工具书.....	61
第二节 检索工具书.....	63
第五章 西文化学化工参考与检索工具书.....	67
第六章 中国化学文献数据库.....	77

机 电 篇

第七章 《工程索引》.....	83
第一节 概况.....	83
第二节 编排体例与著录格式.....	83
第三节 工程主题词表和叙词词典.....	88
第四节 《工程索引》的检索.....	90

第八章 《科学文摘》 ·····	105
第一节 概况·····	105
第二节 编排结构与著录·····	106
第三节 INSPEC 叙词表·····	110
第四节 《SA》的检索·····	114
第九章 《金属文摘》 ·····	122
第一节 概况·····	122
第二节 编排结构与著录格式·····	122
第三节 《冶金词表》与主题词组配·····	126
第四节 《金属文摘》的检索·····	127
第十章 中文机电参考与检索工具书 ·····	133
第一节 参考工具书·····	133
第二节 检索工具书·····	138
<u>综 合 篇</u>	
第十一章 知识产权与专利文献 ·····	143
第一节 知识产权概论·····	143
第二节 专利基础知识·····	145
第三节 专利文献及其检索·····	148
第十二章 标准文献及其检索 ·····	172
第一节 概况·····	172
第二节 标准文献的手工检索·····	175
第三节 标准文献的网络检索·····	177
第十三章 国内综合数据库 ·····	182
主要参考文献 ·····	194

文献检索基础知识

第一节 概 论

科学技术的发展,具有连续性和继承性,科学技术的发明创造,需要依靠经验、材料和理论的不断积累,没有科学上的继承和借鉴,就没有提高。任何一个科技工作者,都有赖于在前人已经取得成就的基础上进行新的研究和探索。科技文献是科学技术研究成果的记录,积累了许多有用的事实、数据、理论、方法和科学假设,反映了科学技术研究的进展和水平,是科技信息的主要来源。

随着科学技术的迅速发展,科技文献呈现出数量庞大,增长速度极快,出版类型复杂,文种繁多,学科内容交叉渗透等特点。要在浩如烟海的文献中迅速、准确,没有重大遗漏地查找与特定课题有关的文献,就必须了解并掌握文献检索的方法。学习本课程的目的,是通过介绍科技文献检索的基础知识和典型检索工具的使用方法,引导科技人员掌握检索科技文献的方法。

一般认为,信息(情报)检索(Information retrieval)包括三个方面的内容:① 数据检索(Data retrieval),② 事实检索(Fact retrieval),③ 文献检索(Document retrieval)。其中数据检索和事实检索是检索包含在文献中的信息,而文献检索实际是书目检索,检索包含所需情报的文献的线索,再根据文献的线索,进一步查找文献,然后筛选出所需要的信息。

一、信息、知识、情报和文献

信息是物质存在的反映,不同的物质具有不同的存在方式和运动规律,从而构成不同的特征,各自发出不同的信息。20世纪中叶以后,信息的本质不断被揭示,并被引入哲学、传播学、情报学和计算机科学等领域。信息普遍存在于自然界、生物界和人类社会中。根据发生源的不同,信息一般可分为自然信息、生物信息、机器信息和人类信息四大类。湖光山色、刮风下雨是自然信息;细胞染色体的遗传基因是生物信息;电讯系统中的电流脉冲讯号、计算机中使用的“0”与“1”的二进制代码是机器信息;人类社会活动中的各种语言、文字、图形符号是人类信息。

知识是人们通过社会实践积累起来的经验,是对客观世界物质形态和运动规律的认识。人们在社会实践中,不断接收客观事物发出的信息,经过人脑的思维加工,逐步认识客观事

物的本质,这是一个由表及里、由浅入深、由感性到理性的认识过程。由此可见,信息是产生知识的原料,知识的产生来源于信息,知识是被人们理解和可认识并经大脑重新组织和系列化了的信息,是可被记录和被物化的劳动产品。

情报是被传递的知识。它是针对一定对象的需要传递的,并且也是生产实践和学科研究中起继承、借鉴或参考作用的知识,是知识的一部分。由此可见,知识要转化为情报,必须经过传递,并为使用者所接受,发挥其使用价值。情报可存在于人脑的记忆中,可存在于事物中,也可存在于文献中。其中,文献是情报存在的主要形式。

文献是记录有知识的一切载体,凡是用文字、图形、符号、声频、视频记录下来,具有存贮和传递知识功能的一切载体都称为文献。由此可见,知识是文献的本质属性,载体是文献的存在形式。

二、文献的级别与类型

按载体的形式,文献可分为:

印刷型:以纸张为载体,以印刷为记录手段而产生的具有悠久历史的一种文献类型,目前仍然是占主导地位的主要形式。

缩微型:以感光材料为载体,以缩微照相为记录手段而产生的一种文献类型,包括缩微胶卷、缩微胶片、缩微平片等。

机读型:以磁性材料和感光材料为载体,通过编码和程序设计,把文字、资料转换成数学语言和机器语言输入计算机,存贮在磁盘、磁带、光盘上的文献。

视听型:以磁性材料和感光材料为载体,以磁记录或光学技术为手段直接记录声音和图像的文献,如唱片、录音带、录像带、电影胶片等。

近年来,随着信息技术的飞速发展,电子出版物大量涌现,使文献、情报、信息这三者之间趋向同一,逐渐淡化了三者在概念上的差别,尤其在国际交往中,情报与信息是同一概念(Information)。因此,目前在国内科技界已倾向于用“信息”一词替代情报。

(一) 文献的级别

按照内容、性质和加工深度不同,文献可分为:

(1) 一次文献(即原始文献)。是文献著者在科学研究、生产实践中,根据科研成果、发明创造撰写的原始文献,如专著、期刊论文、会议论文、专利文献、学位论文等。一次文献的参考和使用价值较高,是文献的主体,是最基本的信息源,是文献检索的主要对象。

(2) 二次文献。是指将分散、无序、查阅不便的一次文献,按照一定的原则进行加工、整理、简化、组织,如著录(即记录)文献的外部特征,摘录内容要点,使之成为便于存贮、检索的系统,如目录、题录、文摘、索引等检索工具的文献。它不仅报道文献内容,更可以提供检索一次文献的线索。

(3) 三次文献。是指在利用二次文献的基础上,选用一次文献的内容,进行分析、概括、评价而产生的文献。它源于一次文献,又高于一次文献,是一种再创性文献,如专题述评、动态综述、教科书、参考工具书、词典、手册等。三次文献一般来说系统性好,综合性强,内容比较成熟,常常附有大量的参考文献,有时可作为查阅文献的起点。

(4) 零次文献。是指未经出版发行的,或未进入社会交流的最原始的文献,如私人笔



记、手稿、试验记录、原始统计数字等,其主要特点是内容独特新颖,但一般不成熟,难以获得,构成为非正式渠道交流的信息源。

(二) 文献的类型

根据性质、特点和出版形式的不同,文献主要可分为十种类型。

1. 图书(Books)

图书大多是对已发表的科技成果、生产技术知识和经验,经过著者的选择、鉴别、核对、组织、加工而成的,图书一般论述比较系统、全面、可靠,代表了某一时期某一学科的发展水平。图书查阅方便(有目次表、索引),但出版周期较长,信息内容更新不够迅速。如果要对范围较广的问题获得一般的知识,或对陌生的领域获得初步的了解或研究探讨某一专业问题,阅读图书是有效的办法。图书一般属三次文献,但有的专著往往包含著者的新观点,或使用的新方法、新材料等,往往具有一次文献的意义。图书种类较多,包括专著(Monograph)、丛书(Series of monograph)、教科书(Textbook)、词典(Dictionary)、手册(Handbook)、百科全书(Encyclopedia)等各种阅读型图书及参考工具书。图书在大型图书馆有广泛的收藏,比较容易获取。

2. 期刊(Periodicals)

期刊又称杂志(Journal, Magazine),一般是指具有固定刊名,定期或不定期出版的连续出版物。期刊着眼于其出版周期短、报道文献速度快、内容新颖、发行及影响面广等特点,能及时反映科技新成果、新水平、新动向,是科技成果比较快捷的公开形式;期刊同时着眼于其丰富的内容特征,拥有众多的作者和读者,因而成为利用率最高的文献类型。期刊发表的论文大多数是原始文献,许多新成果、新观点、新方法往往首先在期刊上刊登。科技人员应熟悉本专业有关的期刊,经常阅读期刊可以了解动态,掌握进展,开阔思路,吸收新的成果。据估计,从期刊上得到的信息约占信息来源的65%以上。期刊论文是文献的主要类型,是检索工具报告的主要对象。国内公共图书馆、高校图书馆和省市以上信息所都有收藏,期刊原文亦比较容易获取。

3. 科技报告(Technical reports)

科技报告是与政府部门签有合同的科研项目的报告,反映的是新型科学和尖端科学的研究成果。其特点是:内容新颖、专业性强,能代表一个国家的研究水平。科技报告的类型有技术报告(Technical reports)、札记(Notes)、论文(papers)、备忘录(Memorandum)、通报(Bulletin)等。科技报告一般单独成册,有单独的篇名、机构名称和统一的连续编号(即报告号)。许多国家的政府机构和大公司企业,为了发展国防工业和尖端科学,为了获得利润,拨出巨额经费用于科学技术研究,这些研究成果一般都要写成科技报告,从而使科技报告成为一种数量庞大、内容复杂的重要情报源。目前,全世界每年都有大量的科技报告产生,估计约有80万件,其中以美国政府研究报告为主。美国政府报告由PB、AD、NASA、DOE四大套报告组成。属于密级的报告,在认为没有必要保密时,便解密成为公开文献。在我国,国家图书馆、国防科技信息研究所和上海图书馆所收藏的科技报告相对比较完整。

4. 政府出版物

政府出版物是各国政府部门及其所属的专门机构发表、出版的文献。就文献的性质来看,其内容可分为行政性文献(如政府公报、法规、方针政策、调查统计资料等)和科技文献(科技报告、科普资料、技术政策等)两大类。政府出版物所包含的内容广泛,从基础科

学、应用科学到政治、经济、军事等社会科学,几乎涉及整个知识领域,但重点在政治、经济、法律、军事等方面。通过这类文献,可了解一个国家的科学技术、经济政策、法令、规章制度等。

5. 会议文献(Conference papers)

会议文献是指在国际学术会议和各国国内重要学术会议上发表的论文和报告。它们都是围绕着某一学科或专业领域的新成就和新课题来进行交流、探讨的,一般都要经过学术机构严格的挑选,代表某学科领域的最新成就,反映该学科领域的最新水平和进展趋势,兼有直接交流和间接交流两种方式的长处。

6. 专利文献(Patents)

专利文献是根据专利法公开的有关发明的文献,主要包括专利说明书,也包括专利法律文件和专利检索工具。

7. 学位论文(Thesis, Dissertation)

学位论文是高等学校、科研机构的研究者为获得学位,在进行科学研究后撰写的学术论文,是经过审查的原始成果。学位论文一般要有全面的文献调查,比较详尽地总结前人的工作和当前的研究水平,作出选题论证,并作系统的实验研究及理论分析,表明自己的观点,探讨的问题往往比较专一,是带有创造性的研究成果,是一种重要的文献来源。我国实行学位制度以来,比较重视对国内学位论文的收集。1984年教育部决定,我国所有研究生的博士、硕士学位论文,一律交中国科技信息研究所收藏,并提供中国学位论文的缩微平片。目前万方数据公司和清华电子出版社均以数据光盘、网络服务和建立镜像站点等形式提供学位论文服务。

8. 技术标准(Technical Standards)

技术标准是对工农业产品、工程建设或其他技术项目的质量、规格、参数、检验方法及技术要求所作的统一规定,是人们在生产、设计和检验过程中共同遵守和使用的技术依据。技术标准按使用范围可分为:国际标准、区域性标准、国家标准,专业标准和企业标准等五大类型。每一种技术标准都有统一的代号和编号独自构成一个体系。技术标准是生产技术活动中经常利用的一种信息源,省级以上的技术监督研究所和科技信息研究所均有所收藏。

9. 科技档案

科技档案是指具体工程建设及科学技术部门在技术活动中形成的技术文件、图纸、图片、原始技术记录等资料,包括任务书、协议书、技术指标和审批文件以及研究计划、方案、大纲和技术措施、技术调查材料(原始记录、分析报告等)、设计计算、试验项目、方案、记录、数据和报告等,还包括设计图纸、工艺和其他有关材料。科技档案是生产建设和科学研究工作中用以积累经验,吸取教训和提高质量的重要文献,大多由各系统、各单位分散收藏,一般具有保密和内部使用的特点。

10. 产品样本

产品样本是国内外生产厂商或经销商为推销产品而印发的企业出版物。主要用来介绍已经生产的产品的品种、特点、性能、结构、原理、用途、维修方法和价格等。查阅、分析产品样本,有助于了解产品的水平、现状和发展动向,获得有关在设计、制造、使用中所需的数据和方法,产品样本常常附有外观照片和结构图,专利产品还注有专利号(根据专利号可查找



专利说明书),对于新产品的设计、试制都有较大的实际参考价值。

第二节 检索工具和检索语言

一、检索工具

检索工具是用来报导、存贮和查找文献线索的工具。它是在文献的基础上,经过加工整理、编辑而成的二次文献,任何检索工具都有存贮和检索两个方面的职能,存贮的广泛、全面和检索的迅速、准确是对文献检索工具的基本要求。检索工具一般须具备四个基本条件:① 必须详细著录文献的外部特征(篇名、著者姓名、文献序号等)和内容特征(文献的学科属性和主题内容等);② 必须具有既定的检索标识(如文献的分类标识、主题词、著者姓名、文献序号等);③ 全部文献必须根据标识,系统地、科学地排列,成为一个有机的整体;④ 能够提供多种检索途径。

检索工具按著录内容可划分为目录、题录、文摘和索引四种。

1. 目录

目录是对图书、期刊或其他单独出版物特征的揭示和报道。目录以单位出版物为著录对象,一般只记录外部特征,如题名、著者、出版事项、载体形态等。目录主要用于检索出版物的名称、著者及其出版、收藏单位。常用的目录有国家书目、馆藏目录(专题目录,联合目录、出版发行目录等)。

2. 题录

题录是对单篇文献外表特征的揭示和报道。著录项目一般有篇名、著者、文献来源、文种等,由于著录项目比较简单,因此收录范围广,报道速度快,是用来查找最新文献的重要工具。报道题录的检索工具的名称不统一,有的定名为目录,有的定名为索引。这是因为题录在形式上与功能上分别与目录、索引相似的缘故,但就性质而言,题录与目录、索引有根本的不同。著名的题录刊物有美国的《化学题录》(Chemical Titles)、《现期期刊目次报道》(Current Contents)、英国的《当代工艺索引》(Current Technology Index)等,我国出版的《全国报刊索引》也属这种类型。

3. 文摘

文摘是系统报道、累积和检索文献的主要工具,是二次文献的核心。文摘以单篇文献为报道单位,不仅著录一次文献的外表特征(即文摘的标头部分,其著录项目与题录相同),还著录文献的内容摘要。按文摘报道的详简程度,文摘可分为指示性文摘和报道性文摘两种类型。指示性文摘亦称简介,内容简单,通常仅介绍文献的论题范围和研究目的;报道性文摘用精练的语言报道原文的主要内容,如讨论的目的范围、采用的方法、重要的公式数据、研究结果和结论等。报道性文摘有时可代替原文,对于不懂原文文种及难以获得原文的科技人员尤为重要。

4. 索引

索引是查找图书、期刊或其他文献的词语、概念、篇名或其他知识单元的检索工具,它通常由一系列按字顺或其他逻辑次序排列的检索标识和文献条目指引符号两部分组成。这些检索标识可以是论文的题目、名词术语、人名、地名,也可以是分子式、各种号码等。索引的

基本功能是揭示文献内容和指引读者查找文献。进行数据检索和事实检索主要应使用各类工具书的索引来完成。

索引是一种附属性的检索工具,它不但广泛应用于各类型的文献中,也广泛应用于各种检索工具之中。一般文摘性检索工具主要由文摘(或题录)和索引两部分组成,文摘(或题录)主要起报道作用,索引主要起检索作用。衡量一种检索工具质量如何,索引系统的完备性是一个重要标志,学习检索工具的使用方法,主要是学习索引的使用方法。

二、检索语言

检索语言,也称标引语言(标引文献特征语言)、索引语言(编制索引的语言),是专门用来描述文献特征和表达检索提问的人工语言,是标引和检索之间的约定语言,它既是编制检索工具的依据,也是检索文献信息的实用语言。检索语言要求接近自然语言,便于检索人员理解和掌握。检索语言必须是单义性语言,一个词只应表达一个概念,一个概念只应用一个词来表示。检索语言的单义性保证了表达概念的惟一性,这就为文献标引和检索提问提供了使用共同语言的基础。检索语言的种类很多,按结构原理可分为分类语言和主题词语言两大类:分类语言主要包括体系分类语言和组面分类语言;主题词语言主要包括标题词语言、单元词语言、关键词语言和叙词语言。按组配方式可分为先组式检索语言和后组式检索语言:先组式检索语言是指表达文献内容的主题词,在检索前已被组配好的检索语言,检索时只能直接利用预先组配好的检索词;后组式检索语言是指表达文献内容的主题词,不预先固定组配,而是在检索时根据检索的实际需要,按照组配规则临时进行组配的检索语言。

(一) 体系分类语言

体系分类语言是按照一定的观点,以科学分类为基础,以文献内容的学科性质为对象,运用概念划分的方法,按照知识门类的逻辑次序,从总到分,从一般到具体,从低级到高级,从简单到复杂,层层划分,逐级展开的一个层累制号码体系。体系分类语言是一种先组式检索语言,用分类号作为检索标识。分类号采用国际上广泛使用的拉丁字母和阿拉伯数字,简洁明了,便于标引和检索。

“类”是指具有许多共同属性事物的集合,体系分类语言就是一种体现知识分类等级制概念的标识系统,反映了事物之间隶属、平行和派生的关系,每划分一次,就形成一批并列的概念(下位概念),使体系分类语言具有按学科或专业集中,系统地揭示文献内容的功能,便于从学科和专业的角度进行全面的检索。如“艺术”类可以划分为“绘画”、“雕塑”、“摄影艺术”等。

体系分类表是标引和检索文献的共同依据,如《国际十进分类法》、《中国图书馆图书分类法》等。

(二) 主题语言

主题语言是一种描述性语言,它用语词直接表达文献的主题,这些语词就是表达主题概念的标识。根据选词原则、组配方式和使用规则等,主题语言又可分为标题词语言、关键词语言和叙词语言。

1. 标题词语言

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语言。标题词是通过对文献内容的分析,从中找出具有实质意义,能表达文献内容特征,并经过规范化处理的词、



词组或短语。

由于标题词语言是先组式检索语言,组配固定,不易混乱,所以标题词表是标题词及其使用规则的汇编,是标引和检索文献的工具,是一部标题词词典。标题词一般包括主标题词和副标题词两级。主标题词表示事物本身概念,一般是事物或过程的名称;副标题词表示事物的各个方面,如性质、部分、方法、操作等名称,主标题与副标题固定组配,按字顺排列。

利用标题词语言编制的检索工具组配固定,不易混乱,还可通过“见”和“参见”扩大检索途径,查全率较高,其缺点是系统性较差,内容相关的文献易分散在多处,不能按标题词的任意组配来进行自由检索。

2. 关键词语言

关键词语言是从文献的标题、摘要或正文中,抽取能表达文献主题、有实质意义的词汇作为标引文献内容特征的标识。关键词语言是不受词表控制,未经规范化处理的自然语言。

由于关键词语言是一种非控制语言,只要指令计算机排除不具有检索意义的非关键词,就可用计算机进行自动抽词标引,编制关键词索引。这些非关键词是:①冠词、代词、连词、感叹词、介词、系动词和某些助动词、形容词;②某些缺乏检索意义的通用性名词,如设计、性质、技术等。

由于关键词语言能适应使用计算机的需要,因而大大加速了文献标引过程,缩短了检索工具出版的时间,便于用来检索最新文献。关键词常常采用轮排的方式编制索引,一篇文献有多少个关键词,就可轮排多少条索引,大大增加了检索点,使用灵活方便。

关键词是未经规范的自然语言,自然语言中的多义词、同义词、近义词造成同一主题文献的分散,查准率和查全率较低。所以,关键词语言仅用来编制临时性的辅助索引,以加快检索刊物的出版速度。CA(美国《化学文摘》)每期文摘后均附有关键词索引,但没有编制卷索引和累积索引。

3. 叙词语言

叙词语言是从文献内容中选取经过规范化的、从概念上不可再分的、具有组配功能的、能表达各学科基本概念的名词术语和指示特定事物的专用名词为叙词,通过叙词以及叙词的相互组配来标引和检索文献。组配是叙词的基本特征,叙词的组配是概念的组配。进行检索时,只要根据检索要求,从叙词表中选取相应的叙词,按组配规则,任意组配检索概念,既能形成泛指性强的检索标识,概括主题所涉及的各有关学科,提高查全率;也能形成专指度高的检索标识,表达主题所包含的复杂概念,进行特性检索,提高查准率;而且还可形成若干个检索标识,进行多途径检索。因此,叙词语言是近代用途较广,既适合于手工检索、更可用于计算机检索的后组式检索语言(叙词是后组式语言,但当一些概念组配不能严格、全面地反映原概念意义时,就直接采用复合概念作为叙词)。

叙词表是以规范化、受控的动态性的叙词为基本成分,并以参照系统显示词关系。用于标引和检索文献时,叙词表是准确选取叙词的工具,因此也称为主题词表,检索词典。

叙词表一般由一个主表和若干辅表组成。主表将全部叙词按字顺排列,每个叙词著录成一条款目,包括表示词之间等同、等级项相关关系的参照项、范畴号、含义注释、族首词、双语对照等;辅表是主表的各种辅助索引,帮助检索主表。辅表一般有:①范畴索引,又称分类索引。它将叙词按概念所属的学科性质,分成若干大类,每个大类再分若干小类,小类下将叙词按字顺排列,形成一个类似体系分类的概念分类系统,便于从学科分类角度查找叙

词。② 词族索引,也称等级索引。该索引利用概念成族原理,将有关的叙词汇集成一族,以概念最广的一个叙词为族首词,构成一个从泛指叙词到专指叙词的等级关系。词族索引按族首词的字顺排列,从族首词就可以找到一系列同族的叙词及其层层隶属关系。词族索引是实现扩检,满足族性检索的重要手段。③ 轮排索引,即将叙词中的复合词按其所包含的单词字顺作轮排,以便于从任意单词迅速查找包含该单词的复合词。④ 双语(或多语)对照索引,提供从另一种(或几种)语言查词的途径,也便于标引外文文献时选取叙词。⑤ 专有叙词索引,即以地区、人物、机构、产品等专有名词为叙词的索引,即把叙词表中上述专有名词的叙词分离出来,另编索引,是主表的组成部分。

叙词表按选词的学科范围分为综合性叙词表和专业性叙词表,我国编制的《汉语主题词表》、《化工汉语主题词表》分别属于这两种类型。

第三节 检索原理、途径、方法和步骤

就高等教育的目的而言,必须培养大学生的自学能力、研究能力、思维能力、表达能力和组织管理能力等,而文献检索技能是培养这些能力的基础。通常,获取文献信息的方法有直接检索和间接检索两种。在浏览文献的过程中,直接获得和累积所需信息,称为直接检索。直接检索可以明确判断文献所包含的情报是否符合需要,科技人员一般习惯通过浏览本专业的核心期刊、专著、会议录以及使用检索工具的有关类目积累有关信息,以了解本专业及有关课题的发展动态,但要数量庞大、高度分散的科技文献中迅速、准确、全面地检索某课题的信息,必须利用检索工具进行系统地查找。利用检索工具获取文献信息的方法,称为间接检索,间接检索是一种科学的检索方法,我们平时所说的文献检索,主要是指间接检索。

一、检索原理

间接检索就是从按照检索标识排列的文献集合中,获取所需文献或文献中所包含的信息内容的查找过程,也就是代表检索需要的检索提问标识与检索系统中存贮的文献特征标识逐一查证比较的过程。要使文献检索能够顺利进行并取得理想效果,关键是存贮文献时选用准确的检索语言。间接检索必须利用检索系统,完整的检索系统包括文献的存贮和文献的检索两部分内容,存贮是检索的基础,检索是存贮的目的。

存贮文献时,首先按照一定规则,把文献的外部特征和内容提要著录下来,形成能反映文献特征的记录单元,称作款目。文献的外表特征包括文献的篇名、著者和出处(刊载文献的出版物的题名、卷期号码、年月、页数等)。这些外表特征可形成文献的外表特征标识,如篇名、著者、号码等。然后,对文献的内容进行分析,按学科属性或主题概念用合适的检索语言(分类符号或规范化的科学名词、词组)进行标引,形成文献的内容特征标识。

文献经过著录和标引,形成了反映文献外表特征和内容特征的各种标识,这些标识分别按照一定的排检规则编排成严格有序的排检序列,输入文献检索系统,为检索文献提供条件。检索文献时,根据检索要求,进行检索提问。检索提问用所需文献的特征标识来表示,如已知所需文献的外部特征,可以将外部特征标识作为检索提问。但在大多数情况下,检索者对所需文献的外部特征不了解,这时就要对检索要求进行内容分析,提取所需文献的内容特征标识,选用准确的检索语言进行标引,从而得到检索提问标识,再将检索提问标识输入



检索系统,与存贮在系统中的文献特征标识进行比较,当检索提问标识查证到与其一致的特征标识时,存贮的文献就被检索出来了。

二、检索途径

检索途径是指通过文献的特征检索文献的途径。通过对文献特征的描述、标引和组织,构成各种目录、索引等检索工具,这些检索工具为检索文献提供了多种途径。

1. 题名途径

以文献题名(书刊名、篇名)为检索标识的检索途径。主要使用《图书书名目录》、《期刊刊名目录》、《报刊论文题名索引》等。这类目录、索引以文献题名的字顺排列。

2. 著者途径

以文献著者(个人与团体著者、编者、译者、专利权人、专利申请人等)姓名和名称为检索标识的检索途径。主要利用著者目录、著者索引、专利权人索引、专利申请人索引等。这类索引通常按个人著者姓名与团体著者名称的字顺混合编排。

3. 号码途径

以文献出版社所编号码或文献收藏者所编号码为标识检索文献的途径。主要利用专利号索引、报告号索引、入藏号索引、标准号索引等。这类索引通常按带大写字母的顺序加号码顺序编排。

4. 分类途径

以文献内容的学科属性在不同分类体系中的分类号为标识检索文献的途径。主要利用图书分类目录、报刊资料分类索引、专利分类索引、标准分类目录等。这类索引按分类号的顺序编排。

5. 主题途径

以代表文献主题内容的主题词为标识检索文献的途径。主要利用主题索引、化学物质索引、关键词索引等。这类索引以主题词(标题词、关键词、叙词等)的字顺编排。分子式索引是以化学物质的另一种形式——分子式为标识检索文献,也属主题途径的类型。

前3种途径是利用文献的外表特征为标识检索文献的途径,需要特定的已知条件;后两种途径是从文献内容的学科属性或主题概念检索文献,使用极为广泛。

三、检索方法

1. 常用法

利用检索工具检索文献的方法,又可分为顺查法、倒查法和抽查法。

顺查法。以课题提出的年代为起点,由远及近,逐年查找,查到最近。用这种方法查得的文献比较完整,能反映课题研究发展的全过程,查全率高,但工作量大。

倒查法。由近及远,按时间逆序查找。这种方法的重点在近期文献,近期文献不仅反映最新水平,而且一般都引用、概括了早期文献。这种方法省时灵活,效率较高,但易产生漏检。

抽查法。对课题研究进展最快、发表文献最多的若干年的文献逐年查找。这种方法省时、效率高。

2. 追溯法

利用最新发表的文献(最好是切合课题的综述性文章)后面所附的参考文献为线索,由

近及远追溯查找,也可从了解到的著者姓名利用检索工具查找。这种方法是不少科技人员常用的方法,方便实用,但亦会产生漏检。

3. 循环法

又称分段法,是常用法和追溯法相结合的方法。先利用检索工具查出一批有关文献,再利用这些文献所附的参考文献和原文中涉及的线索追溯查找,这样分段交替进行,直到取得满意的效果。

四、检索步骤

文献检索的过程是根据研究课题的要求,选择检索系统,确定检索标识,按照一定的检索途径和方法查找特定文献的过程。因此,文献检索一般分为以下5个步骤。

(1) 分析课题:① 掌握与课题有关的某些知识、名词术语以及课题所属的学科范围或技术领域等内容;② 分析课题的研究目的,明确检索要求;③ 分析课题直接或间接提供的已知条件。

(2) 选择检索系统,确定检索标识:① 根据课题的学科性质、所需的文献类型选择合适的检索系统;② 根据所具备的条件选择手工检索工具或计算机检索数据库,也可采用二者结合的方法;③ 选择报道及时、收录文献全面、索引系统完备的检索系统;④ 既要选择使用综合性的检索工具,也应注意选择使用专业性的,以及单一文献类型的检索工具。检索系统确定后,根据不同检索系统的要求,利用主题词表、分类表、索引指南等标引核对检索标识。

(3) 确定检索途径和检索方法:① 检索途径的选择要从检索要求、已知条件和文献检索工具的结构等几个方面综合考虑。题名途径、号码途径一般用来查找某篇特指的文献;著者途径用来查找某著者(或某学术团体、企业)一定时期内的工作动态;而要系统检索某一课题的文献,应主要从分类途径和主题途径入手。由于大部分检索系统不编制分类索引(专利索引除外),仅是大的部分按分类编排,且类目较粗,因而只适用于最新文献的浏览,因此,主题途径是最重要、应用最为普遍的途径,它既适用于回溯性检索,也适用于检索最新文献。② 检索方法的确定还要根据所提供的检索工具、检索人员对主题背景材料掌握的深度以及使用检索工具的熟练程度而定。

(4) 查找文献线索:按照确定的检索标识和检索途径,利用检索工具的索引查找文摘号,根据文摘号查找文摘或题录。在查找过程中,要根据查找的具体情况不断分析、调整检索标识、检索途径和检索方法,直至达到满意的效果。

(5) 查找和获取原始文献:对检索的文献线索进行研究和筛选,如需原始文献,可根据文摘、题录等所提供的文献出处,向文献收藏单位索取原始文献。

计算机检索基础知识

第一节 计算机检索概况

一、计算机检索发展简史

人类进入 20 世纪后,科技文献资料的数量不断高速增长,使得文献处理、信息查询成为繁琐的劳动,于是人们开始寻找新的解决办法和手段。穿孔卡片检索装置的问世,标志着人类进入了机器检索时代,此后发展起来的缩微胶卷检索装置,使机器检索在小范围内推广开来。早期的机器检索的功能和效率都不太理想。40 年代中期,人们发明了电子计算机,这就为高速而有效地查找文献提供了努力的方向。随着计算机技术和数据库技术的逐步成熟,文献检索迎来了现代化的时代,到了 20 世纪末,互联网络的建设,使计算机检索渐渐走入人们的生活。

二、计算机检索系统的类型

按信息存贮与检索使用的设备可把计算机检索系统分成以下几种类型。

1. 脱机检索系统

这是早期的一种系统,国外在 20 世纪 60 年代达到实用水平。当时,美国化学文摘社(CAS)等一些大型书本式检索工具的出版单位开始出版机读磁带版,人们在这基础上建立了脱机检索系统。它的工作原理是:用户把检索要求提交给检索者,检索者把它们编制成提问文档存放在穿孔纸带或磁带上,计算机读入该提问文档后,顺序地扫描机读磁带,找到与之匹配的文献。由于用户与系统之间不能实时对话,响应速度也慢,因此常常是将一批提问汇集起来成批检索,所以也叫批式检索。80 年代初,我国许多科研机构 and 大学引进了这种技术,现在已被新的检索系统所取代。

2. 计算机辅助缩微式检索系统

自 20 世纪 20 年代起,缩微技术就得到了长足的发展。缩微制品具有存贮量大、占地小、成本低、保存期长等优点,目前仍有大量的论文、技术报告、专利说明书、报纸、手稿等文献,以缩微制品的方式保存着。缩微胶片需要借助专门的阅读器才能阅读,而现在的阅读器都带有手动或自动检索装置,能从大量的缩微画面中找出符合需要的画面,这就是缩微检索系统。计算机辅助缩微式检索系统则是把计算机的倒排文档检索技术和数据库检索技术引入到缩微检索系统中,可以实行人机对话,由计算机来检索和启动阅读器。

3. 联机检索系统

20 世纪 70 年代,联机检索系统投入商业运营。它是由大型计算机系统、几十至几百种联机数据库以及一批检索终端组成。联机检索系统可以实行人机对话和远程联机检索,到目前为止仍然是主要的信息检索工具之一。它的工作原理是:用户在自己的终端机上通过通讯线路与系统的主机连接,用户输入检索提问式,系统主机进行查找并把结果反馈给用户,用户可以随时对检索式进行调整。它的缺点是要支付昂贵的通讯费用。

4. 光盘检索系统

20 世纪 80 年代,光盘技术发展成熟并被大量使用。人们把数据库存贮在容量大、价格低廉的光盘上,借助光盘驱动器和微机读取检索数据库。单机光盘检索系统随着光盘数据库的不断增长,查找一个课题往往需要频繁更换光盘,于是人们利用光盘塔或大硬盘构建了光盘检索网络系统。它是以微机的局域网为基础,拥有可同时运行多张光盘的光盘塔驱动器,同时提供给多用户使用同一张光盘或不同光盘的网络软件。由于光盘检索系统造价较低,在我国高校图书馆得到普遍使用。

5. 视频数据系统

视频数据系统也叫电视信息系统,20 世纪 80 年代在发达国家进入普及阶段。它是以用户的直观显示器或经过改装的彩色电视机作为终端,直接接收电视中心播放的信息或与视频数据系统中心的数据库进行联机对话,存取检索信息的。我们通常所说的图文电视就是视频数据系统的一种。视频数据系统在信息产业被广泛使用,如证券中心、中国内贸部信息中心、商品交易所等。

6. 国际互联网 Internet

Internet 是全球最大的计算机网络,它联接近千万台主机,拥有上亿的用户,是目前最大的信息网络资源。互联网就像一个虚拟的社会,有网上商店、网上医院、网上图书馆、网上新闻等等。网上的信息似乎无穷无尽,没有秩序,但借助于搜索引擎,人们可以查找到自己所需的站点,这是网络检索的主要内容之一。另外,一些数据库生产单位也纷纷建立网上数据库,比如联机检索系统的代表——Dialog 系统。借助互联网来传输数据要比使用电话线经济得多,随着家用电脑上网的日渐普及,互联网络检索将成为人们的必修课程。

第二节 数据库及检索

计算机检索中应用最广的是数据库检索。本节我们将对数据库检索的基本知识作一简单介绍。

一、数据库的一般结构

不同的数据库,其结构都不一样,提供的检索途径也各不相同,但它们的基本结构却具有共同的特征。

(一) 字段、记录、文档

数据库最基本的特征是它们都由文档组成,而文档又由若干记录组成,组成记录的则是字段。

字段是文献著录的基本单元,每个字段反映了文献的一项特征。在数据库中每个字段



往往都有一个字段名,如:论文作者的字段名为 AU,论文题目的字段名为 TI 等等。

记录是由若干不同字段组成的文献单元,一个记录一般代表一篇文献。在数据库中每个记录都有一个记录号,其意义类似于检索工具中的文摘号。

文档是由若干数量的记录所构成的数据集合。它主要应用于大型联机检索系统中,如 DIALOG 系统中 399 文档是美国化学文摘,221 文档是世界专利索引,在检索中输入不同的文档号就能对不同的数据库进行检索。

(二) 文档结构

记录排列顺序不同,则文档结构就不一样,而这将对检索速度造成直接影响。

简单文档,是按照记录输入先后顺序存放数据记录的文档。它不考虑记录之间的关系。其优点是增加记录简单,但修改、删除、检索记录则很费时间,数据库一般都不采用这种结构。

顺序文档,其记录按照某关键字的值依次顺序排放。该关键字是指其值能惟一标识该记录的字段,比如记录号。顺序文档中记录的检索、修改、删除都比较简单,可使用折半查找法提高速度,但增加记录比较麻烦,要插到适当的位置才能保证文档的顺序。

索引文档是指文档外另带了个索引表,索引表中的每项内容包括一个关键字的值和对应于该值的相应地址。一般,索引文档本身记录的物理顺序和逻辑顺序可以一致,也可以不一致。而索引表是按照关键字值的大小顺序排列的,当索引表本身十分庞大时,还要建立索引的索引,形成多级索引。索引文档是常见的文件存取结构之一。

倒排文档,是将顺序文档中某些代表文献特征的字段抽出来,再按一定的规律重新排列而成的文档。它的出现,满足了人们从文献的其他属性值进行查找的需要,大大提高了检索效率,因而它也是绝大多数数据库采用的文档结构。其缺点是要占用额外的存贮空间,在主文档进行更新时,倒排文档的更新也要相应进行。

二、数据库的检索

(一) 检索的一般步骤



我们进行检索前,肯定会有一个检索需求,因此,首先要明确自己的检索目的,然后就此选择数据库。不同的数据库其学科范围、收录文献的类型都不一样。选择好数据库后,就要确定检索词,编制检索式。等检索结果显示出来后,我们可以进一步判断是否准确,要不要进一步限制或扩大检索范围。最后把命中的文献打印出来。

(二) 检索词

检索词包括两大类:受控词和非受控词。受控词是一种事先规范好的语言,它们取自主题词表、叙词表、分类词等。用受控词作为检索词,其检索效率高,在检索过程中若数据库提供了词表,则可优先使用。非受控词即自由词,是用户在检索过程中习惯使用的词汇,其缺点是可能与数据库本身抽取的词不符,造成漏检。

检索结果不尽如人意的原因大致有以下几点: