

文献检索 及其自动化基础

郭金亭 胡锡成 编

WENXIANJIAN
SUOJIQIZI
DONGHUAJICHU

南海出版公司

655

7

《文献检索及其自动化基础》编委会

主 编	郭金亭	胡锦涛	汤利光
副主编	陈东韵	王振兴	
编 委	(以姓氏笔画为序)		
	王学军	王振兴	龙训田
	汤利光	陆宝益	周 萍
	杨兴华	胡锦涛	秦立宽
	郭金亭	郭宽民	高荣菊

前 言

文献检索作为一门方法课,对充分利用现存的文献资料,使自己少走弯路,提高社会和经济效益,特别是对于当代大学生,面对当前信息量的迅猛增长,掌握文献检索的基本原理和方法,在信息和文献资料的海洋中自如地选择自己需要的信息,永远立于不败之地,更具有重要的意义。

本书以原国家教委关于“文献检索教学基本要求”的精神为指导,对内容和结构进行了合理的安排,以适应不同学校及不同专业的教学要求。本书在编写过程中注意文献检索理论体系的建设,全面系统地讲述了文献检索的基本理论及各类工具书的类型与特点,强调参考工具书的利用。在选材上,充分注重各校学生的专业特点,注意典型性,以增加其实用性。在讲授理论的基础上,考虑到各专业的特点,比较全面地阐述了文献检索的基本方法和基本途径,以便培养学生检索与利用文献的能力。特别是针对当前计算机技术的迅速发展及其在文献管理、信息(数据)传递方面的应用,机检已成为必然趋向之时,适时地将自动化检索的内容引入本书,并专设两章,详细阐述了这一方面的内容。在教学过程中建议根据各校的具体情况,适当增加机检内容,以培养学生的感性认识,尽快适应目前高新技术快速发展的形势。另外,在每一章之后,均附有一定数量的思考与练习题,便于学生实习。有少量综合性实习题具有一定难度,可在教师指导下进行检索实习,以达到综合利用基本原理,熟悉各种专业检索与参考工具书的目的。

本书的内容设置和体系结构比较完整,适用范围比较广,可作

为各类院校开设文献检索课的教材。根据各校特点建议安排 30 课时,授课重点可结合各校特点灵活掌握,授课课时与实习课时比以 2:1 为宜。由于各馆目前所使用的应用软件不同,本书中所介绍的一例未必与各校的具体情况相同,可根据各自的情况进行补充。为更好地培养学生的学习兴趣,特别是有关 Internet 的内容,建议尽量增加学生的上机机会,并开通 Internet 网,为学生学习创造良好的环境。

由于时间较紧,编者水平有限,作者虽对本书的内容设置、结构安排进行了深入的研究,但不妥之处在所难免。诚请各位同行、专家和广大读者批评指正。

编 者

1999 年 7 月

目 录

第一章 文献检索基础	1
第一节 信息、知识、文献和情报	2
第二节 文献概述	6
第三节 文献检索概述	12
思考与练习	23
第二章 检索语言	24
第一节 检索语言概述	24
第二节 分类检索语言	25
第三节 主题检索语言	33
思考与练习	45
第三章 检索工具	46
第一节 检索工具概述	46
第二节 题录性检索工具	50
第三节 文摘性检索工具	53
第四节 索引	59
第五节 中文科技文献检索工具	70
第六节 国外文献检索工具	75
思考与练习	94
第四章 参考工具书	95
第一节 参考工具书概述	95
第二节 参考工具书排检法	96

第三节	参考工具书的类型与特点·····	104
第四节	参考工具书的检索与利用·····	112
	思考与练习·····	180
第五章	图书馆自动化与文献检索·····	181
第一节	图书馆自动化基础·····	181
第二节	图书馆自动化管理软件简介·····	184
第三节	光盘数据库的使用·····	188
	思考与练习·····	198
第六章	网上数据与文献检索·····	199
第一节	网络简介·····	199
第二节	信息传递与检索·····	210
第三节	Internet 漫游·····	230
	思考与练习·····	242
	重要关键词和知识点索引·····	243
	参考文献·····	249

第一章 文献检索基础

随着科学技术的不断发展,特别是20世纪50年代以来,新技术革命的浪潮席卷全球。目前,由于计算机技术和信息技术的飞速发展,Internet网已延伸到世界的每一个角落。只要有一台微机,一个Modem(调制解调器),你就可以“周游世界”,在几分钟之内,访问不同的国度。尤其是文献信息资源电子化程度逐步提高,传统的出版印刷业的状况已得到彻底改变,出版一本文献资料,有时不需要大量人力、物力的投入,只须坐在家里或办公室里、坐在计算机前就可以把文献资料发到任何一位需要此文献的读者手中。于是大量的文献信息犹如洪水般在网络上、市场上、在其他任何可以到达的地方膨胀和泛滥。有用的、无用的、完整的、不完整的,统统涌来。有时人们毫无思想准备或无法找到自己所需要的资料,真正的“信息爆炸”的时代已经来临。

如何在文献和信息的洪流中更全面、更准确地找到自己所需要的文献资料,已是一个急需解决的问题。不解决这个问题,人们的工作或生产效率、成本就不知要增加多少倍,会多走弯路,浪费人力、物力。而文献检索则是解决这一问题的金钥匙。

因此,学习和掌握这把钥匙,已经成为当代大学生的必由之路。

第一节 信息、知识、文献和情报

一、信息、知识、文献和情报的概念

1. 信息

信息论创始人之一的维纳说，信息就是信息，不是物质也不是能量；信息被看成是物质的一种属性，它反映物质客体及其相互作用、相互联系过程中表现出来的状态和特征内容。比如在这一过程中，所发出的消息、信号和信号中所包含的指令、数据及其他内容。由此可以看出：一切事物包括自然界和人类社会都会产生信息。自1948年以来，人们对“信息”的描述或称为定义有几十种之多，但是归纳起来，从科学性和概括性来讲，我们认为信息是被反映物的属性再现。也就是说，信息不是事物的本身，而是由事物发出的消息、指令、数据、信号等所包含的内容。同时，一个完整的信息过程，包括传递、接收、储存、加工和利用等几个方面。

现代信息已是一个非常广泛的社会性概念，它是人类社会共享的一切知识、学问，以及从客观现象中提炼出来的各种消息的总和。它具有如下几个特征：可传递性；可加工性；可存储性；可识别性；可再生性；可共享性；可消除不确定性和可替代性。

事实已经表明，客观世界的三大要素是物质、能量和信息，现代社会中信息已成为人类物质和能源之外的第三资源。人类经历了6000年的农业社会和近300年的工业社会，正逐步向第三个文明社会——即信息社会(Information age)过渡。在信息社会里，信息和知识的占有量已成为社会生产力、竞争力和经济发展成败的决定性因素。因此，掌握信息的收集、加工和利用的方式和方法，对于科技工作者、企业的决策者、教育工作者乃至各行各业都起着十分重要的作用。

2. 知识

人们在实践中将接收到的信息进一步加工处理,分析浓缩,使其发生质的变化而成为知识。知识是认识产生飞跃的结果,是人们在改造客观世界的实践中所获得认识和经验的总和。

知识的产生、发展与利用是一个不断更新的过程。知识本身始终是处在动态的发展过程中,它永远处于不断被充实、完善和更新的状态。

3. 文献

“文献”一词在古代是指典籍及熟悉掌故的人(贤者),始见于《论语·八佾》:“夏礼吾能言之,杞不足征也;殷礼吾能言之,宋不足征也;文献不足故也。足,则吾能征之也”。宋代朱熹在《四书章句集注》中对“文献”一词解释为:“文,典籍也;献,贤也”。现代则是指具有一定历史价值的图书资料和具有某种学术价值的专著、论文等资料。随着社会的发展和科学文化的进步,文献所指的范围也进一步扩大。《中华人民共和国国家标准 BG3792.1-83 文献著录总则》中将文献定义为:“文献:记录知识的一切载体”。由此定义可以认为,古代刻有文字及符号的甲骨直到现代的磁带、胶片、软盘和光盘等均可称为文献。一般详细地将文献定义为:“文献乃是指用文字、图形、符号、声频等技术手段记录人类知识的一种载体,或称为固化在一定物质载体上的知识”。

由文献的定义可以看出,构成文献要有两个条件:第一是要有知识内容。没有任何知识内容的物质不能称为文献。第二是要有用以记录知识的物质载体。载体是信息和知识赖以存在的外壳,没有记录在物质载体上的知识不能称为文献。文献总是记录有一定的信息与知识,总是依赖一定的人工附载物的实体形态而存在。同时,信息与知识被固化在一定载体上以后,便于传播和积累,为人们开辟了更为广阔的交流渠道。

文献具有几个重要特征:信息与知识性;记录性;传播性;积累

性；人工附载物的实体形态性。这五个特征是统一不可分割的，但其中的信息与知识性、人工附载物的实体形态性和记录性是文献的基本特征；传播性和积累性则体现了科学文化的发展与继承。

4. 情报

关于情报的定义，众说纷纭。前苏联情报学家 А·и·米哈依诺夫认为，“情报——作为存储、传递和转换的对象的知识”；英国的 B·C·布鲁克斯则认为，“情报是使人原有的知识结构发生变化的那一小部分知识”；我国著名科学家钱学森说，“情报就是为了解决一个特定的问题所需要的知识”。这些说法的核心都强调了知识的运用。因此归结起来，可以认为“情报是指知识的传递并起作用的部分”。将知识中的特定部分传递到需要这种知识的使用者那里，被传递的这部分知识就成了情报。“传递”是情报的基本属性，也是辨别知识是否转换成情报的重要标志。情报是知识的一部分，具有知识才需要传递。因此，情报具有知识性、传递性和实用性三个属性。

二、信息、知识、文献和情报的关系

信息和知识被记录下来就成为文献，文献是我们利用的主要对象。文献中的知识不可能全部转化为情报，转化的只是特定的一小部分，这部分是读者原来不知道的又能解决实际问题的知识。传递的知识对于传递的对象必须具有针对性、新颖性和实用性。

关于信息、知识、文献和情报之间的范畴关系如图 1—1 所示。

其中：①是尚未被认识或加工的信息区；②是未被传递和记载的知识区；③是非实物型载体的情报区；④是未被利用或未转变为情报的文献区；⑤是“情报 + 知识 + 文献 + 情报”的统一区。

同时，信息、知识、文献和情报是不停地互相转换着的，其转换过程及途径如图 1—2 所示。

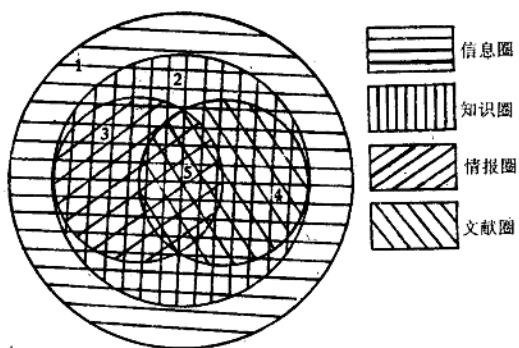


图 1-1 信息、知识、文献和情报的范畴关系

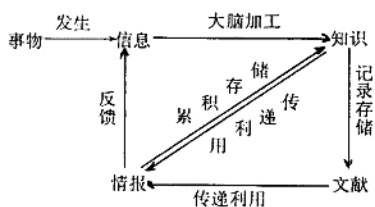


图 1-2 信息、知识、文献、情报的相互转换过程

第二节 文献概述

人们在实践活动中所获得的知识,通过一定载体记录下来就形成了文献。文献不仅记录了当时社会科学文化的状况,为人们进行科学研究提供了丰富的资料,同时推动了社会的进步,是科学文化发展的基石,它在整个社会科学文化及经济发展的过程中起着重要作用。总起来讲,它具有如下几种重要功能:

存储功能。文献既是知识的载体,同时也是情报的载体,它能够存储各个时代的科学文化遗产,是人类存储并供人类分享利用的知识宝库。

传递功能。人们在科学活动的同时,吸取前人的科技文化成果,在吸收时主要是查阅现存的文献资料,即把有用的知识信息传递给了后者。同时,文献是人类传递科技情报的主要媒体。一般科技情报的传递分为正式传递渠道(在刊物上发表或出版专著)和非正式传递渠道(讨论、参观、讲演、展览、通信、交换等手段),前者是主要的传递手段,占情报传递的2/3。

再生功能。人们在研究某一课题时,首先要借助于已有的文献,从中获取大量的间接经验,以便使研究少走弯路,缩短研究周期,提高效率。同时,研究的新成果被记录下来又成了新的文献。产生的新文献从水平上超过了原有文献,所以,这是一个阶梯式发展的、推陈出新的过程。

鉴别功能。科技文献往往是用来衡量科研成果质量高低的重要依据。在评价一项科技成果时,首先要与前人发表的有关文献进行比较,从已有文献提出的观点、方法、工艺和数据等方面逐项核实,以确定该成果的新颖性、先进性和经济效益。

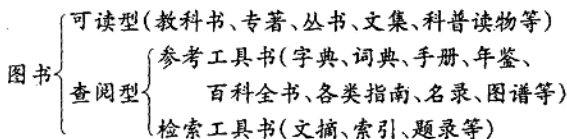
一、文献的类型和特点

从不同的角度,可以将文献分为几种不同的类型。

1. 按文献的内容特点和出版形式划分

按文献的内容特点和出版形式,可以将文献分为图书、期刊、报纸和特种文献 4 种类型。

(1)图书:图书(Book)是用于全面而系统地阐述某一方面或学科的科学理论或发展,传授科学技术以及用来查阅某些知识的文献形式,是最早出现的文献类型之一。根据它的功能还可以分为:



图书的优点是内容一般比较成熟,代表了某一时期和某一学科的发展水平。缺点是出版周期较长。

(2)期刊:期刊(Periodical; Journal; Magazine)是定期或不定期连续出版发行的一种文献形式,主要用于报道某一方面或学科的最新成果、研究状况和其他信息。它是随着科学技术的发展、满足各类信息的快速交流而产生和发展的,是一种非常重要的情报源。按其所报道的范围,可分为综合性期刊和专业性期刊;按其功能可分为报道性期刊(主要报道和刊载研究论文及其他信息等)和检索性期刊(用于文献的检索,是一类重要的检索工具);按出版周期还可分为周刊、旬刊、半月刊、月刊、双月刊、季刊、半年刊、年刊等。

目前,全世界期刊的数量迅猛增加,出版的期刊种类繁多。在我国,通过邮电部门正式发行的期刊就达 4000 余种,非正式发行的刊物还有数千种,而且数量还在增加。网上电子期刊的发行,更使期刊种数剧增。

期刊最大的优点是出版周期短,报道及时。网上期刊则出版

周期更短、时差更小,便于读者了解某一学科或领域的最新研究状况,是读者最常用的一种文献类型。

(3)报纸:报纸(Newspaper)是报道各类最新信息的一种重要文献形式,是一个国家某一地区或部门发布信息及各种消息的重要渠道。它比期刊报道更迅速,周期更短,但报纸报道的信息老化快,范围广但深度不够。对于报纸中的重要内容,将在一定时期后以图书的形式积累下来。

按出版周期,报纸可分为日报、周报、半月报、旬报等。

(4)特种文献:特种文献是指一类特殊形式的文献。主要有:专利文献、会议录、政府报告、学位论文、标准文献、产品说明书等。

2. 按内容特征及加工层次划分

按内容特征及加工层次,可以将文献分为一次文献、二次文献和三次文献。

(1)一次文献:也称为原始文献,是指由科研人员撰写发表的研究成果或记录生产实践活动的文献。它具有内容详尽、使用价值高等优点,是科研人员追踪的主要目标。

(2)二次文献:虽然一次文献具有真实、具体、参考使用价值高等优点,但是它具有比较分散、数量庞大、查阅不便等缺点,并且还有自然语言及其他方面的障碍。而且科研人员的精力有限,不可能全部阅读所有一次文献,致使不能有效地进行利用。为了解决这一矛盾,情报工作者对有情报价值的一次文献进行提炼、浓缩和加工,即著录其外部特征和内容特征,标引出文献的主题,编制成具有多种检索途径的检索工具,即二次文献(如文摘、索引、题录等)。二次文献可使科研人员在较短的时间内有效地检索到所需文献,并保证有较高的查全率和查准率。

(3)三次文献:三次文献则是由某学科或专业的专家对该学科或专业的某一问题的一次文献和二次文献,通过分析综合、浓缩加工而形成的一种文献,如评论、综述、各类参考工具书等。

将文献加工成不同的等级,使繁杂的资料成为有组织、有系统的整体,满足了读者对不同内容和加工层次文献的需求,并且使读者在检索时有章可循,有径可问,同时为加强各种文献和情报的管理和流通创造了条件。

3. 按载体形式划分

为了有效地记录、积累和传递各种情报,人类先后发明和运用了各种各样的物质材料记录科学知识,如刻在甲骨上,铸在青铜器上,焙烧在陶土上,写在简策、布帛和纱纸上;后来印在纸张上,晒在蓝图上,摄在感光材料上,录在磁性材料上,刻录在光盘上等等。总之,随着社会的发展和科学技术的进步,使得文献的载体形式越来越多样化。但总起来讲,目前文献按载体可分为如下5类:

(1)印刷型文献:这是以纸张为存储介质,以印刷(油印、石印、胶印、铅印等)为记录手段而产生的具有悠久历史的一种文献形式。其优点是:便于直接阅读,不用特殊的器材和设备。其缺点是:存储密度低、篇幅大、笨重、占用空间大、整理与保存需要花费较多的人力、物力;文献的识别与提取难于实现机械化和自动化(即难于实现自动化输入与自动化检索)。但它仍是今后较长时间内文献的主体。

(2)缩微型文献:这是以感光材料为存储介质,缩微照相为记录手段而生产的文献形式。主要包括缩微胶卷、缩微胶片等。其优点是存储密度高(一般缩微率从20:1到40:1,随着激光和全息照相技术的应用,缩小倍率最高已达22500;一张全息胶片可存储20万页文献),体积小,节省空间,便于收藏,保存期长,不易损坏变质,价格比较便宜,提取与传递方便,便于实现管理的自动化。其缺点是不能直接阅读,阅读时需要借助于缩微阅读机,容易使读者产生疲劳。另外,不能像纸质文献那样可同时利用几种文献随时加以比较,或对同一文献前后各部分随时进行对比,更不能在文献上作必要的标记或批注。

(3)机读型文献:即计算机可读型文献。它是以磁性材料为存储介质,打字穿孔或光学字符识别装置(即以光电转换和电磁转换)为记录手段,并通过电子计算机处理而形成的一种文献形式,其载体形式主要是磁带、磁盘、光盘等。它主要是通过编码和设计程序,将文献变成数学语言和机器语言,输入计算机,存储在磁带、磁盘及光盘上,阅读时再由计算机将它输出。其优点是存储密度高,存取速度快,可根据需要来组织其体系结构,可以方便地抹去或改变、更新原有记录。缺点是有设备才能阅读,费用较高。但随着计算机技术的逐步发展,以光盘为存储介质的文献正在急剧增加,并为人们所喜爱和接受,且已用于大型的检索工具或参考工具,取得了较好的社会效益和经济效益。

(4)声像型文献:声像型文献又称为视听资料或直感资料。它是以磁性材料和感光材料为存储介质,借助特殊的机械装置,直接记录声音或图像信息而形成的一种文献形式。主要包括:唱片、录音带、录像带、幻灯片、电影片等。其优点是生动直观,能帮助人们认识某些复杂或罕见的自然现象,探索物质结构和运动机制等方面难以用文字来描述的科技情报。它是传播科技知识和科技情报非常有力的工具。

(5)网上出版物:所谓网上出版物是指以电子形态在网络上发布的文献形式,如一些出版社在网络上发行的电子版、Web 浏览页面及其链接(即 Web 资源)等。它具有出版迅速、阅读方便等优点。但由于网络上发行的电子出版物,大多没有经过专人进行鉴定或校对,或由个人随意编辑发布,因而其准确性、完整性和稳定性等方面较差,所以在使用或引用时应加以筛选和斟酌。

二、文献的发展趋势与特点

文献的发展是和社会经济文化的发展密不可分的。社会稳定,经济发展快,则文献发展迅速。进入 20 世纪以来,由于科学技术的进步,各种文献呈现出新的发展趋势与特点。

社科文献出版速度加快,老化较慢,使用寿命往往较长,出版类型多样化。

由于世界经济和科学技术的发展,科技文献的数量激剧增长,造成了科技文献的出版率与科技人员对它的兴趣和利用之间的不平衡,导致研究精力的巨大浪费,出现了所谓的“情报爆炸”。出版类型多样化,内容交叉重复,学科愈分愈细,刊物报道的范围越来越窄。由于稿件急剧增长,致使积压严重。其次,由于科学技术从发现或发明到应用,直到被新的发现或发明所代替的周期越来越短,从而使科技文献的使用寿命越来越短。据估计,各类科技文献的平均寿命约为:科技图书 10~20 年,科技报告 10 年,学位论文 5~7 年,期刊和连续出版物 3~5 年,标准文献 5 年,产品样本 5 年。

另外,科学论文的分布呈现出特殊的规律性。科学论文既集中又分散地登载在大量期刊上,在任何特定的学科或领域中,大部分论文集中在较少量的期刊中。这就是期刊分布的离散状态。这一规律是由英国图书馆学家布拉德福(S. L. Bradford)于 1934 年总结发表的,也称为布拉德福定律(Bradford's Law)。用布拉德福的原话来说,“对于某一主题而言,将科学期刊按刊登相关论文减少的顺序排列时,可以划分出对该主题最有贡献的核心区,以及含有与该区论文数量相同的几个区域。每个区域里的期刊数量呈 $1:n:n^2\dots$ ”。布拉德福当时所做的统计实验结果见表 1-1。

表 1-1 布拉德福的统计实验结果

分区	期刊中载相关论文的数量(篇/年)	应用地球物理学		润滑学	
		期刊数量	相关论文数	期刊数量	相关论文数量
1	>4	9	429	8	110
2	1-4	59	499	29	130
3	<1	258	404	127	152

其中 1 区是核心区。当三个分区的期刊所载相关论文数量的