

XINXI JIANSUO JIAOCHENG

信息检索 教程

姜效先 主编
吴美二
刘二灿

大英百科全书
环境文摘
中国专利文摘
中国大百科全书
化工文摘

中国物资出版社

信息检索教程

姜效先 吴美健 刘二灿 主编

中国物资出版社

图书在版编目(CIP)数据

信息检索教程/姜效先 吴美健 刘二灿主编. —北京:中国物资出版社,

1999.1

ISBN - 7 - 5047 - 1506 - 9

I. 信… II. 姜… III. 情报检索—教材 IV. G354

中国版本图书馆 CIP 数据核字(98)第 23599 号

中国物资出版社出版

北京市西城区月坛北街 25 号 邮编:100834

全国各地新华书店经销

郑州市中兴印刷厂印刷

开本 787mm × 1092mm 1/16 印张 15.0 字数:347 千字

1999 年 1 月第 1 版 1999 年 1 月第 1 次印刷

书号:ISBN - 7 - 5047 - 1506 - 9/TP·0036

印数:0001 - 1500 册

定价:24.80 元

《信息检索教程》编委会

主 编 姜效先 吴美健 刘二灿

副主编 杜 明 陶 钢 王 文 郭美莉

编 委 (以姓氏笔画为序)

王 文 冯有利 刘苏新 李凌杰

杜 明 吴美健 郭美莉 姜效先

陶 钢 夏英华 段殿生

前 言

随着信息化社会的到来,计算机技术的迅猛发展,信息高速公路的不断建设,许多文献检索工具的改革,以及社会对人才需求的变化,使原来的文献检索教材已不能适应形势的发展和需要。为使文献信息检索课更好地培养学生的信息意识,提高他们的综合素质和能力,特邀请多年从事文献检索课教学工作的教师编写此书。

本书全面介绍了文献信息检索基础理论和基本概念,国内外的主要文献检索工具,专利文献,标准文献的检索,详细地阐述了计算机信息检索的基本知识、信息检索技术、检索策略,还用较多篇幅介绍了国际互连网络和光盘信息检索,新增加了书目信息检索。在附录里我们整理了 Internet 网络资源,国外联机检索系统命令以及国内外主要数据库。

本书共分 11 章。各章节执笔者为:第 1 章:吴美健、陶钢、刘二灿;第 2 章:郭美莉、刘苏新;第 3 章:王文、夏英华、杜明;第 4 章:杜明;第 5 章:陶钢、郭美莉;第 6 章:吴美健、杜明;第 7 章:李凌杰、郭美莉;第 8 章:刘苏新、陶钢;第 9 章:冯有利、陶钢;第 10 章:刘二灿、吴美健、王文;第 11 章:刘二灿、段殿生;附录:姜效先、陶钢、王文、刘苏新。

全部稿件经吴美健、刘二灿、陶钢、杜明、王文、郭美莉分别修改后,由姜效先统改定稿。

在编写过程中,参考了近年来最新的文献检索教材和文献工作者的研究成果,在此向他们表示诚挚地谢意。

本书虽经著者努力,但错误、缺点在所难免,希望广大读者批评指正。

目 录

第 1 章 文献信息检索概论	1
1.1 文献、情报、知识和信息	1
1.2 科技文献的系统结构与类型	2
1.3 现代科技文献的整体特征与规律	7
1.4 文献信息检索的目的和意义	9
1.5 信息检索语言	11
1.6 文献检索的步骤和方法	22
第 2 章 文献检索工具	29
2.1 文献检索工具的意义与基本职能	29
2.2 文献检索工具的主要要求	29
2.3 文献检索工具的类型	30
2.4 文献检索工具的一般结构	33
2.5 中文文献检索工具	33
2.6 我国科技文献检索刊物的体系	34
2.7 国内科技文献检索工具的使用	38
第 3 章 国外科技文献检索工具	41
3.1 美国《工程索引》	41
3.2 英国《科学文摘》	49
3.3 FAO《农业索引》	55
3.4 美国《环境文摘》	60
3.5 日本《科学技术文献速报》	65
3.6 美国《数学评论》	71
第 4 章 科技报告及其检索	80
4.1 科技报告概况	80
4.2 美国的科技报告	81
4.3 美国科技报告的检索	82
4.4 中国的科技报告	84
4.5 国内外其他科技报告检索工具要目	84
第 5 章 专利文献及其检索	86
5.1 专利文献	86

5.2	中国专利文献及其检索	92
5.3	英国《世界专利索引》及其检索	96
第6章	标准文献及其检索	104
6.1	标准文献概述	104
6.2	中国标准文献	105
6.3	国际标准文献	108
6.4	美国标准	111
6.5	英国标准	112
6.6	联邦德国(前)标准	113
6.7	日本工业标准	114
6.8	前苏联(现俄罗斯)标准	115
6.9	法国标准	117
第7章	数据库类型	119
7.1	文献数据库	119
7.2	非文献型数据库	125
第8章	计算机信息检索技术	132
8.1	计算机信息检索发展过程	132
8.2	计算机信息检索技术	132
8.3	文档信息检索技术	140
8.4	中文全文信息检索系统	142
第9章	国际联机信息检索	144
9.1	国际联机检索原理与服务方式	144
9.2	几种国际联机检索系统简介	145
9.3	Internet 网络信息检索	145
第10章	光盘信息检索	153
10.1	光盘检索概述	153
10.2	《中文科技期刊光盘数据库》检索	161
10.3	《中国学术期刊(光盘版)》检索	164
10.4	万方数据库检索	170
10.5	《中国专利文摘》光盘检索	177
10.6	其他光盘数据库简介	183
第11章	书目信息检索	186
11.1	OCLC 信息检索	186

11.2 北京图书馆“文津”文献管理系统	189
11.3 深圳图书馆自动化集成系统(ILAS)	191
11.4 中国国家书目只读光盘	195
11.5 丹诚图书馆集成系统	196
11.6 MILINS 管理系统	197
11.7 博菲特文献管理集成系统	199
附录 1 国外主要联机检索系统命令一览表	202
附录 2 ESA 系统数据库简介	204
附录 3 ORBIT 系统数据库简介	210
附录 4 常用国际联机数据库	213
附录 5 我国自建的数据库	215
附录 6 Internet 网络资源	217
主要参考文献	221

第 1 章 文献信息检索概论

1.1 文献、情报、知识和信息

1.1.1 文献

在我国,“文献”一词始见于《论语·八佾》:“夏礼吾能言之,杞不足徵也;殷礼吾能言之,宋不足向徵也;文献不足故也。足,则吾能言之。”大意是,孔子能讲解夏、殷的典章制度,但杞、宋两国的典章制度因缺乏足够的文献而无法证实。对孔子所说的文献,宋代朱熹在《四书章句集注》中注释为:“文,典籍也,献,贤也。”这就是说,文献包括典籍和宿贤。典籍指有关典章制度的文字资料。宿贤指阅历丰富、满腹经纶的贤人。

到了现代,由于科学技术的发展,出现了各种各样的载体材料,发明了各种各样记录知识信息的方式,文献大量涌现,文献概念的外延也在不断扩大。按照国际标准化组织《文献情报术语国际标准》(ISO/DIS5217)对文献的定义即是:“在存储、检索、利用或传递记录信息的过程中,可作为一个单元处理的、在载体内、载体上或依附载体而存储的信息或数据的载体。”不仅包括了传统的书刊、文稿等,而且包括了缩微品、音像资料、机读资料和电子出版物等新型的信息知识载体。所以我们可以将文献理解为:记录有信息或知识的一切载体。

1.1.2 情报

在我国 1979 年以前出版的“辞源”和“辞海”中,对于情报的解释是“战时关于敌情之报告,曰情报”。在 1979 年出版的“辞海”中,扩充为“……亦泛指一切最新情况报道。如科学技术情报”。这就是说,“情”,即情况、消息;“报”,即报道、传递、交流。因此,情报就是情况、消息的报道与交流。

1.1.3 信息、情报、知识和信息化

N.维纳指出“信息是我们适应外部世界并且在使这种适应为外部世界所感到的过程中,同外部世界进行交换的内容的名称。”这是从通讯的角度,考虑机器和动物体内部的通讯,再转到动物体之间的通讯给信息下定义的。信息是由符号和信号来传递的。

现在,信息的概念范围较广,包括人与人,人与计算机,动物界和植物界等等的消息交换。在当今社会,信息无处不在,人们几乎时刻也离不开信息。

“Information”作为广义理解,可译成“信息”,作为狭义理解,被译成“情报”。因此情报是狭义的“信息”。

英国情报学家 B.C.布鲁克斯认为:“情报是使人原有的知识结构发生变化的那一小部分知识。”

苏联情报学家 A.H. 米哈依洛夫认为情报是：“情报——作为存储、传递和转换的对象的知识。”

日本《情报组织概论》一书认为：“情报是人与人之间传播着的一切符号系列化的知识。”从以上定义来看，情报与知识的概念紧密地联系在一起，知识和传递是情报的基本属性，知识是人们对客观对象规律性的认识。从科学概念和客观对象的规律性来考虑，信息、情报和知识之间的逻辑关系为包含和被包含的关系。如图 1-1 所示。

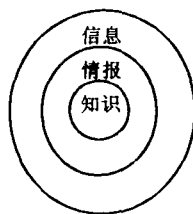


图 1-1 信息、情报、知识关系

信息化，简单地说是指国民经济从发展以物质和能源为基础向以知识和信息为基础的转变过程。或者说是指国民经济的结构框架重心从物理性空间向知识性空间转变的过程。主要表现在：

(1) 信息部门规模

它在国民生产总值中占的比重，发达国家一般占 40~60%，我国 1985 年为 15%。

(2) 信息工作的基本装备水平

主要指计算机、电话机和电视机等。

(3) 信息对生产要素的增值作用

指信息物化为生产力而产生的经济效益。

我国信息处于发展阶段，情报信息尚未网络化和系统化。在科学技术革命瞬息万变，新技术层出不穷的新形势下，我们面临的是一个信息化的社会。钱学森特别指出：“信息既是新技术革命的核心内容，也是迎接新技术革命最核心的对策。将来的社会是信息化的，而信息是一种重要的资本、成本的核心，经济上的重要资源。”

社会需要情报信息的主要原因：

(1) 信息网络一般不受地区、部门等隶属关系的限制

由于信息横向流通，适应了发展商品经济所必须的开放式全方位的需要，因而地区、部门的限制较少。

(2) 信息网络是实现信息资源共享的好形式

少花钱多办事，提高了设备、资料利用率，经济效益好。在经济杠杆起作用的单位和部门，在互利协作原则下，信息网络得到了发展。

(3) 信息网络是实现情报专业化的条件，是情报社会化的必然过程

为了向用户提供有深度的权威情报产品及情报服务，情报信息工作必然走向专业化，社会化。

(4) 信息网络是一个世界性的发展趋势

1.2 科技文献的系统结构与类型

1.2.1 科技文献的系统结构

就科技文献整体而言，其内容按层次可分为：一次文献、二次文献和三次文献。其结构如图 1-2 所示。

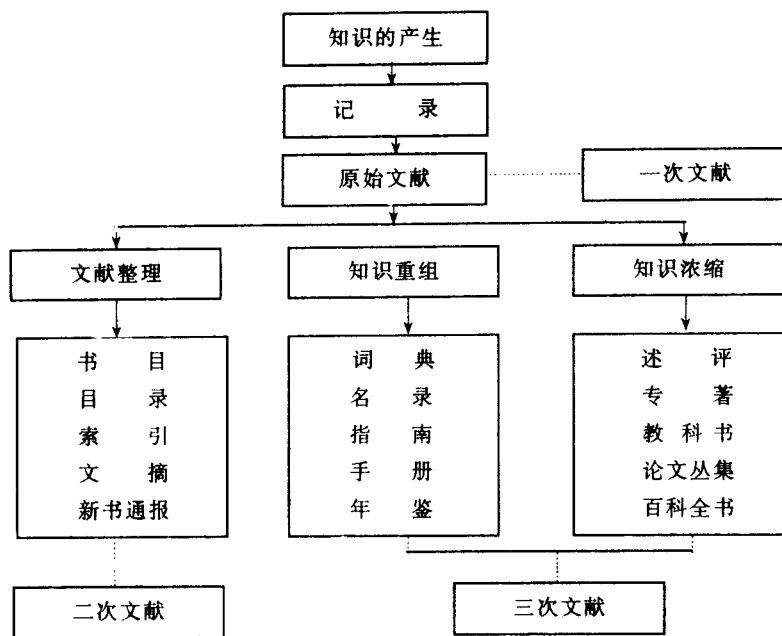


图 1-2 科技文献系统结构图

一次文献(Primary Document)通常是指原始制作,即作者以本人的研究成果为基本素材而创作(或撰写)的文献。如期刊论文、专利说明书、会议论文、科技报告和学位论文等等。此外,它还包括一些不公开发表的文献,如实验记录、日记、备忘录、手稿、内部报告、技术档案、信件等。

二次文献(Secondary Document)是指文献工作者对一次文献进行加工整理后所得到的产物,也是为了便于管理和利用一次文献而编辑、出版和积累起来的工具性的文献。它包括书目、索引、文摘等。二次文献的重要性在于可以帮助人们查找一次文献。

三次文献(Tertiary Document)是指利用二次文献,选用一次文献内容而编纂出的成果。如词典、手册、年鉴、百科全书、专著、教科书、论文丛集、述评、文献指南以及书目之书目等等。

科技文献结构示意图表明,从一次文献到二次文献、三次文献是一个由博而约、由分散到集中、由无组织到系统化的过程。也是科技文献的层次结构由无序走向有序、由一种有序结构演变成另一种完善的有序结构的过程。

了解了科技文献的系统结构,就可以使情报检索有的放矢,减少盲目性,情报传递与情报检索存在着互逆关系。在检索时一般是通过三次文献了解二次文献,查到有关的文献信息,或者通过二次文献获得原始文献的线索,进而尽快地在原始文献中查到所需的知识或信息。

1.2.2 科技文献的类型

1. 按科技文献的物质形式划分

(1)印刷型文献,这是以纸张为存储介质,以手写、印刷、打字、复印等为记录手段而产生的一种传统的文献形式。其优点是便于阅览和流传,不受时间、地点和条件的限制;其缺点是存储密度太低,体积庞大笨重,占储藏空间过多,不易管理和保存,难于实现自动输入和自动检索。

(2)缩微型文献,这是一种以感光材料为存储介质,以缩微照像为记录手段而产生出来的文献形式,包括缩微胶卷、缩微胶片、缩微卡片。其优点是:体积小,存储密度大,传递方便,节省储藏空间,一张全息胶片存储 20 万页纸张文献;其缺点是不能直接阅读,阅读者须借助于缩微阅读机。

(3)机读型文献,这是一种以磁性材料,如磁盘、磁带或磁鼓等为存储介质,以打字、穿孔或光学字符识别装置为记录手段,并通过计算机处理而产生出的一种文献形式。其优点是存储密度高,存储速度快,原有记录可更改;其缺点是需要先进的技术设备才能阅读,使用费用高。

(4)声像型文献,这是一种以磁性材料或感光材料为记录介质,以电磁转换或光电转换技术为记录手段直接记录声音、图像而产生的一种文献形式,也叫视听或直感资料。如唱片、录音带、录像带、科技电影、幻灯片等。其优点是存储密度高,直观,真实,可直接表现那些难以用文字描述的事物,其缺点是费用高。

(5)光盘型文献,这是一种用激光束记录和再现信息的存储装置。它作为知识或信息的一种新载体,自 70 年代以来,随着信息自动化处理技术,特别是微自动化处理技术的发展,为知识信息高密度存储展示了广泛的前景。光盘可分为两种类型,一种是信息编码模拟式录像光盘,又叫激光电视唱片,是比较理想的图像存储载体,它可以存储和再现固定或活动的图像。另一种是通过数字化(0,1)处理数据的数字光盘,这种将信息数字化的光盘,英美等国称为 CD(Compact Disc 高密度光盘),它是通过激光束照射在光盘上的槽或泡的变形来写入或读出信息。光盘型文献的优点是存储密度高,使用寿命长,读写速度快,存取时间短,是当前最新的科技文献载体物质。光盘型文献是最新的一种文献形式,由于它是通过计算机读写,所以也是机读型文献的一种。

在以上 5 种物质形式的科技文献中,缩微型科技文献、机读型科技文献、声像型科技文献和光盘型科技文献等几种新型载体的文献正在迅速发展,数量不断增多,比重日益增大,大有与传统的印刷型科技文献相抗衡的趋势。但是印刷型科技文献在众多文献载体中仍占首要地位,最受读者欢迎,利用率最高,而且在今后相当长一段时间内仍是各文献部门的收藏重点。

2. 按科技文献出版形式划分

(1)科技图书

科技图书范围广,包括专著、文集、教科书、普及读物、百科全书、年鉴、手册、词典等。其特点是:内容系统、全面、成熟、可靠。有的图书往往还包含著作者本人的新资料、论点和方法。如果想对范围较广的问题获得一般知识,或对陌生的问题获得初步了解,参考科技图书是十分有效的。科技图书的发行量逐年增加。

(2)科技期刊

科技期刊是一种周期性出版物,具有品种多,数量大,出版周期短,报道速度快,内容新颖,能及时反映当前科技水平等特点。期刊文献多数是一次文献,是人们传递科技信息,交流学术思想所使用的最基本、最广泛的手段。

根据美国对科学家的调查表明,情报需求的 68% 是期刊论文。英国电气工程师所用情报源中 70% 是专业期刊,据统计,从期刊方面获取的科技信息约占整个信息来源的 65%。期刊是科技人员吸取成果,掌握进展,了解动态,开阔思路的重要参考文献。

(3) 科技报告

科技报告是关于某项科研成果的正式报告,或者是对研究过程中阶段进展情况的实际记录,它的特点是每份报告自成一册,有连续编号,内容专深,由于它是研究的记录和成果,因而代表着一个国家和专业的科研水平,对科研工作能起直接借鉴作用。

科技报告基本上都是一次文献(少数书目索引也被编入科技报告)。许多最新的研究课题与尖端学科的资料,往往首先反映在科技报告中,现在全世界每年大约发表 70 多万件科技报告,占世界文献总数的 10% 左右。

(4) 会议文献

会议文献是指在学术会议上宣读的论文或交流的书面论文。它分为会前文献和会后文献两种。会前文献包括会议预印本,会议论文摘要,会议议程和发言提要,会议近期通报或预告等。会后文献包括会议记录、会议专刊、技术报告等。

由于会议日益增多,科技会议文献大量产生,目前国际上每年出版的各种专业会议录近 3000 种,据统计多数科技会议每次提出的论文 9~50 篇,多者数以百计,最多可达千篇以上。

科技会议文献的特点是:出版形式不固定。同一个会议的文献论题集中,内容新颖、丰富、专深、学术性强,能反映某学科或某专业的当前状况,往往代表着一门学科或某个专业的最新成果,反映着国内外科学技术的最新发展水平和趋势。所以,它是了解各国科技发展水平和动向的重要科技文献。

(5) 政府出版物

这是各国政府部门及其设立的专门机构发表、出版的文件,内容广泛,大致可分为行政性文件(如法令、统计等)和科技文献。其中科技文献占整个政府出版物的 30%~40% 左右,包括政府所属各部门的科技研究报告、科普资料和技术政策等文献资料。它们在未列入政府出版物之前,往往已被所在单位出版过。因此,它与其他科技文献(如科技报告等)有重复,但也有的是初次发表的。目前,许多国家都设有专门机构(如美国政府出版局、英国皇家出版局等)负责办理政府出版物的出版发行工作。据不完全统计,美、英、法、日等国的政府出版物每年多至几万种,并且还在逐步增加。

政府出版物集中反映了各国政府各部门对有关工作的观点、方针、政策,对了解某一个国家的科学技术和经济状况及政策,具有一定的参考价值。

(6) 专利文献

专利是用法律来保护科学技术发明创造所有权的制度。当专利申请案提出后和批准时,一般就公布由发明人呈交的说明该项发明的目的、技术梗概和专利权限的申请说明书和正式说明书——这就是所谓的专利文献。

(7) 标准文献

它主要是对工农业产品和工程建设的质量、规格及其检验方法等方面所作的技术规定,

是从事生产、建设的一种共同技术依据。每一件技术标准都是独立、完整的资料。它作为一种规章性的技术文献,有一定的法律约束力。对标准化对象描述的详尽性、完整性和可靠性,绝非一般的杂志论文、样本、专利所能比拟的。

(8)学位论文

学位论文即高等学校研究生、本科生毕业时写作的作为评定学位的论文。由于它一般不出版,只供应复制品,取得的手续也较麻烦,因而不易为读者所利用。中国科学技术情报研究所收藏有部分复制品。学位论文质量参差不齐,所探讨的问题较专,有时在某些方面有独到的见解,对研究工作有一定的参考价值。

学位论文是非卖品,不发行,但也有印成单行本,或在期刊上发表摘要的,少数也有全文发表的。

(9)产品样本

产品样本是制造厂商为了推销产品而发出的以介绍产品性能为主的出版物。如产品目录、产品说明书、产品总览、简明手册等。其内容大致为已定型产品的性能、用途、结构原理、使用方法、操作规程、产品规格。根据其内容和出版形式的不同,分为以下几种类型:单项产品样本、产品说明书、产品目录、企业产品一览、产品样本集、产品数据手册、厂商企业介绍、厂刊、外贸刊物等。

其特点是:技术成熟可靠,图文并茂,形象直观,内容全面具体,出版迅速,发行范围广泛,装璜美观,设计新颖,重宣传推广,轻信息,新陈代谢快。产品样本具有鲜明商业性质,但因大多数样本附有产品性能、规格、外形照片、结构简图和线路图等,所以,它对科技人员选型和设计,为进口国外产品和设备具有参考价值。

(10)科技档案

科技档案是科学技术研究和生产建设部门,在科学技术研究和生产建设活动中形成的有具体工程对象的技术文件,图样、图表、图片、原始记录的原本或代替原本的复制本。它包括任务书、协议书、技术经济指标和审批文件、研究计划、方案、大纲和技术措施;有关的技术调查材料(原始记录、分析报告等)、设计计算、试验项目、照片、影片、录像、记录、数据和报告;设计图纸、工艺卡片及应该归档的其他材料等。

技术档案的特点:1)反映本单位科学技术研究、生产建设活动真实的历史记录,其内容真实、详细、具体、准确可靠。2)数量庞大,是科技文献的最完善、最可靠的形式。3)它保存期限较长,是科技文献的一种主要形式。而现代科学技术继承性很强,在科研工作中采用现有的技术成果,是多快好省搞科研的捷径。科学技术储备越雄厚,起步就越高,科技发展也愈快。所以,对有价值的科技档案要永久保存。4)科技档案保密性较强,一般都有密级限制,借阅手续严格。

科技档案的种类多,也很复杂,一般有工程设计档案、基本建设档案、生产技术档案、设备档案、科学研究技术档案等。它是继续进行科研工作和生产建设的重要依据;是科学研究和历史研究的必要条件;是进行科技交流的重要工具。科技档案是科学研究和生产建设工作中积累经验、吸取教训和提高质量的重要科技文献。

1.3 现代科技文献的整体特征与规律

现代科学技术的进步日新月异,无论是从它的发展速度、发展规模来看,还是从它对人类社会和经济生活的影响来看都是前所未有的。作为记录和传播科学信息或知识的文献,更是直接受到科学技术发展的巨大影响。

当今科学发展的最显著特点是“高速度”与“综合性”。“高速度”首先表现在科研成果的大量涌现、科学知识的急剧增加。由于科学研究领域日益宽广,科研队伍日益扩大,科研事业集约化程度日益提高,因而科学研究的新成果大量涌现。其次,这种高速度还表现在科学知识迅速地转化为社会生产的周期。19世纪平均有50年,今天则只要10年。“综合性”表现在各学科间的渗透、交叉、转移、组合日益加强,一系列边缘科学、综合科学、交叉科学、横断科学相继出现,整个科学构成一个网络式的立体结构。重大科研课题往往涉及众多的学科,它的完成必须依靠多学科的协作攻关。

当代科学发展的上述特点必然反映到文献中来,使现代文献从整体上呈现以下特征。

1.3.1 数量急剧增长

科学技术的迅速发展,各种知识门类的不断增加,无疑会导致各个知识领域的文献数量急剧增长。国外统计资料表明,科技成果每增加1倍,情报量就增加几倍;生产量翻一番,文献情报量就增加4倍。

几十年来,许多文献学家都在研究、探寻文献的增长规律。美国文献学家D.普赖斯(D. Price)统计了期刊在世界范围内两个世纪的增长情况。他以科学期刊量为纵轴,以历史年代为横轴,把各不同年代的科学文献量在坐标图上逐点描述出来,然后以光滑曲线连接各点,得出了表示文献增长的曲线图——普赖斯曲线图。它表明了科学文献与时间成指数函数增长的规律。

文献数量庞大,增长迅速,说明文献资源丰富。但数量浩繁的文献也产生了“信息污染”,使图书信息机构在选择、收集、整理、保存、传递文献方面,面临许多新的课题。

1.3.2 内容交叉重复

现代社会文献量爆炸性增长,与文献的冗余规律有密切关系。现代科学技术综合交叉、彼此渗透的特点,必然导致知识的产生和文献的出版也相互交叉,彼此重复,具体表现为:

(1)各种学术机构、研究单位在科研选题上相互重复,反映其研究成果的文献内容也必然出现重复。

(2)同一内容的文献以不同的形式、不同文字发表或出版。

(3)世界各国为了及时了解和利用其他国家的科技成就,相互翻译出版了大量的书刊资料。

(4)再版和改版的文献数量在增多。科学技术的迅速发展、知识更新速度的加快,使原来的文献内容变得陈旧。为了及时反映当前科技发展的水平,作者需要不断地修正、充实自己原来的著作,于是引起再版和改版文献的大量出现。

(5)许多杂志社同时出版内容完全相同的印刷型和缩微型两种版本。

(6)各国出版商为提高声誉或追求盈利,大量出版发行热门书和新兴学科书刊。

冗余文献虽然能扩大人们获得和接触这些文献的机率,但也使识别和使用这些文献变得非常困难。有资料表明,各类文献中,有用的信息内容仅占25%。这就要求我们在收集文献时,必须加以认真区分,筛选,以避免不必要的重复,造成浪费。

1.3.3 类型、文种复杂

现代文献的生产突破了传统的纸张印刷方式,声、光、电、磁等现代技术和化学塑胶新材料的广泛应用,使现代文献载体形式发生了重大变化,缩微资料、声像资料、机读资料、光盘资料等新型文献载体相继问世。这些非纸质文献载体,或加大了知识信息的储存密度,或加快了信息的检索、传递的速度,或使人闻其声,见其形,获得直观的感受。总之,从功能上大大优于传统的纸张印刷型文献。因此,这些新型的文献载体在文献中所占的比例越来越大。综合收藏与使用多种载体文献类型,是现代图书馆的一个明显特点。

世界各国的文献使用的语种也在不断增多。据联合国教科文组织统计,现在全世界出版的科技文献中,有不少于一半的文献是用50%以上的科学家不懂的语言发表的。语言障碍已成为文献收集、整理和利用的严重问题。

1.3.4 载文聚散有序

现代科学技术不断分化综合的发展趋势,使各学科严格界限逐渐消失,各学科之间的相互联系逐渐加强。由于这一原因,使得文献的分布呈现出既集中又分散的不均匀现象,相当数量的专业论文相对集中刊载在少量的专业期刊中,其余数量的专业论文却高度分散刊载在大量非专业期刊中。

文献分布的不均匀现象还表现在:一种专业期刊不仅刊载本学科的论文,也发表许多相关学科或相邻学科的论文,而同一专业的论文不仅发表在本专业的刊物上,也出现在许多不同专业的刊物上。

英国化学家、文献学家布拉德福(S. C. Bradford)经过长期对各学科文献的大量统计调查,发现了文献分布规律。他发现,全部有关电技术的文献约1/3登载在本专业少数几种期刊上,约1/3登载在数量约5倍的并非直接与电有关的交通运输等相关学科的期刊中,还有1/3的有关电技术的文献,登载在25倍数量的相邻学科期刊上。布拉德福在对书目、文摘等进行大量统计分析的基础上,采用等级排列技术,揭示了文献离散定律。他指出:“如果把科学期刊按其关于某一学科的文章刊载的数量多少,以渐减顺序排列起来,在所得的清单中,可以分出直接为此学科服务的期刊所形成的核心,和另外几个组或区,其中每组或每区期刊所载的文章数量,同核心区中的期刊刊载的文章数量相等。这时,核心区中的期刊数量与相邻各区中期刊的数量成 $1:n:n^2$ 的关系”。

布氏定律表明,某一学科文献在期刊上载文量的多少,是随着该期刊与该学科的疏密程度发生增减变化的。关系越密切,载文量越多,期刊的种数就越少;关系越疏远,载文量越少,期刊的种数就越多。按专业文献载文量多少,可以将期刊划分为3个区域,每一区域中期刊登载某一学科文献数量,是该学科所发表文献总数量的1/3,而3个区域的期刊数量之比成几何级数分布。其中,第一个区域为核心区,是载文量最高的少数几种核心期刊。第二区域为相关区域,是载文量中等的数量较多的期刊。第三区域为相邻区域,是载文量最低而

数量最多的期刊。布拉德福在 1948 年提出的文献聚散经验公式是：

$$P_1:P_2:P_3=1:n:n^2$$

P 代表不同区域期刊种数, n 代表布拉德福常数, 按已分析的数据, n 的数值约为 5。

布氏定理表明, 每一学科或专业的文献, 在科技期刊中的分布, 总是相对集中在少数专业期刊中, 同时又高度分散在数量庞大的相关专业与相邻专业的期刊中。专业核心区期刊, 种数不多, 本学科文献载文率高, 信息量大, 与本学科关系最密切; 相关区期刊, 种数较多, 本学科载文率中等, 信息量次之, 与本学科关系较密切; 非专业相邻期刊, 种数很多, 本学科载文率低, 信息量小, 与本学科关系较疏远。一般来说, 核心期刊的载文率必须在 50% 以上, 而且读者的借阅率高, 引用指数较高, 从量与质两方面测定才比较合理。

1.3.5 新陈代谢频繁

科学技术的迅速发展, 新理论、新观点、新技术、新产品的层出不穷并迅速更新, 加速了知识与信息的新陈代谢; 记录知识与信息的文献的有效使用时间日益缩短, 失效周期日益加快。旧的文献被新的文献所代替后, 其使用的读者越来越少, 使用频率越来越低。文献的有效使用期称为文献的寿命, 文献的失效周期称为文献更新期。文献寿命的衰减, 就是文献老化, 文献更新的开始。文献的寿命的长短, 表现在不同文献类型的有效使用期与不同学科文献的老化失效期方面, 是以读者使用时间作为判断标准的。各种类型的文献, 有不同的有效使用时间。

1.4 文献信息检索的目的和意义

1.4.1 文献信息检索的定义

信息检索(Information Retrieval)是指将信息按一定的方式组织和储存起来, 并根据信息用户的需要找出有关信息的过程。所以, 它的全称又叫信息存储与检索(Information Storage and Retrieval), 这是广义的信息检索。狭义的信息检索则仅指该过程的后半部分, 即从信息集合中找出所需信息的过程, 相当于人们所说的信息查寻(Information Search)。

作为检索对象的信息, 它有不同形式, 有的以文献形式出现, 有的以数据和事实出现。根据检索对象形式的不同, 信息检索又分为文献检索、事实检索和数据检索。

(1) 文献检索(Document Retrieval), 这是以文献(包括文摘、题录和全文)为检索对象的一种检索。凡是查找某一主题、某一时代、某一地区、某一著者、某一文种的有关文献, 以及回答这些文献的出处和收藏处所等, 均属于文献检索的范畴, 文献检索是一种相关性的检索, 检索结果是文献线索, 还须进一步找到这些文献, 阅读后, 才能决定取舍。

(2) 数据检索(Date Retrieval), 以数据为对象的一种检索, 包括文献中的某一数据、公式、图表, 以及某一物质的化学分子式等。数据检索分为数值型与非数值型, 它是一种确定性检索。检索结果只是有与无、正确与错误之分, 它是直接可以利用的情报。

(3) 事实检索(Fact Retrieval), 以事实为对象, 查找某一事物发生的时间、地点及过程。事实检索与数据检索一样, 是一种确定性的检索。因此, 有人认为事实检索只是数据检索中的一个方面。