

信息检索 与利用

冯凯 王筱明 主编



华东理工大学出版社

EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

信息检索与利用

冯 凯 王筱明 主编



华东理工大学出版社

图书在版编目(CIP)数据

信息检索与利用/冯凯,王筱明主编. —上海:华东理工大学出版社,2005.7

ISBN 7-5628-1711-1

I. 信... II. 冯... ②王... III. 情报检索-高等学校-教材 IV. G252.7

中国版本图书馆 CIP 数据核字(2005)第 044689 号

信息检索与利用

.....

主 编/冯凯 王筱明

责任编辑/李国平

封面设计/王晓迪

责任校对/张 波

出版发行/华东理工大学出版社

地 址:上海市梅陇路 130 号,200237

电 话:(021)64250306(营销部)

传 真:(021)64252707

网 址:press.ecust.edu.cn

印 刷/江苏省通州市印刷总厂有限公司

开 本/787×1092 1/16

印 张/15.5

字 数/392 千字

版 次/2005 年 8 月第 1 版

印 次/2005 年 8 月第 1 次

印 数/1-4050 册

书 号/ISBN 7-5628-1711-1/TP·133

定 价/23.50 元

前 言

自 20 世纪下半叶以来,人类产生的信息量高速增长,据估计,这期间,全世界每年出版图书 50 多万种、期刊 10 万多种、专利约 50 万件、科技报告约 90 万件、会议文献 10 多万篇、产品样本 50 多万种,每年发表的科技论文总数近 500 万篇,并呈指数式增长,真可谓信息浩如烟海。如何从这浩如烟海的信息中找出所需信息,就成为信息检索教学的重要任务。为此,原国家教委于 1984 年发出过文件,要求在高校开设“文献检索与利用”课,多年来,在文献检索教学尤其是手工文献检索教学方面已取得很大成绩。然而,自 20 世纪 80 年代以来,以缩微平片、声像带、磁盘、光盘等形式记录的非纸质信息急剧上升,伴随着计算机进入多媒体时代,信息科技也步入多媒体发展时期,手工检索靠“手翻、眼看、大脑判断”的检索方式已难以全面适应当今信息的发展,计算机信息检索必然地提到了应用与发展阶段,以 Internet 为代表的全球性网络的实际应用更进一步推动了这一发展,这既是对手工检索的扩展,也是时代的要求。

信息检索(Information Retrieval),是指将信息按一定的方式组织和存储起来,并根据信息用户的需要找出有关的信息过程,所以它的全称又叫“信息的存储与检索”(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要的信息的过程,相当于人们通常所说的信息查寻(Information Search)。

计算机信息检索,是指以计算机技术为手段,通过光盘和联机等现代检索方式进行信息检索的方法。

与手工检索一样,计算机信息检索应作为未来科技人员的一项基本功,这一能力的训练和培养对科技人员适应未来社会和跨世纪科研都极其重要,一个善于从电子信息系统中获取信息的科研人员,必定比不具备这一能力的人有更多的成功机会,故大家对这一领域的技术应予以足够重视。

许多大学都开设了“信息检索与利用”的课程,部分学校还专门为研究生开设了这门课,这足以看出大学教育对“信息检索与利用”课程的重视。本教材主要为大学本科生的教学而编写。在教学过程中,大家可以根据专业性质的不同在内容上进行适当的删简和取舍。

本书由长期从事信息检索教学的教师结合多年教学经验编写而成,其中第一章、第二章由王筱明编写;第三章、第九章由冯凯编写;第四章由胡光林编写;第五章由周鸣镝编写;第六章由胡春编写;第七章、第八章由危志明编写。

由于编者水平有限,本书可能出现疏漏和错误,恳请广大读者批评指正。

编 者

2005 年 6 月

目 录

第一章 信息检索的基本知识	(1)
第一节 信息的概念	(1)
第二节 文献的定义	(1)
第三节 文献的特点	(1)
第四节 文献的特征	(3)
第五节 文献的级别、种类和载体形式	(4)
第二章 信息检索原理	(9)
第一节 信息检索原理、类型和方法	(9)
第二节 信息检索工具	(11)
第三节 信息检索语言	(12)
第四节 信息检索的方法和途径	(14)
第三章 中文检索工具及其使用	(16)
第一节 中文检索工具概述及结构体系	(16)
第二节 中文检索工具编排及其检索途径	(18)
第三节 国内科技文献检索工具简介	(20)
第四章 国外检索工具与使用	(32)
第一节 美国《化学文摘》(CA)及其检索方法	(32)
第二节 美国《地质学题录与索引》(BIG)及其检索	(55)
第三节 美国《工程索引》(Ei)及其检索	(69)
第四节 美国《科学引文索引》(SCI)及其检索方法	(84)
第五章 数据与事实检索	(103)
第一节 数据与事实检索概述.....	(103)
第二节 工具书概述.....	(103)
第三节 常用工具书简介.....	(105)
第六章 古籍资料的检索	(132)
第一节 书目的利用.....	(132)
第二节 类书的利用.....	(135)
第三节 政书的利用.....	(138)
第四节 索引的利用.....	(142)

第七章 计算机检索基础	(152)
第一节 计算机检索概述.....	(152)
第二节 检索技术.....	(154)
第三节 检索的基本程序.....	(156)
第四节 常用搜索引擎.....	(158)
第八章 国内外数据库检索	(161)
第一节 万方数据库的使用简介.....	(161)
第二节 重庆维普数据库的使用简介.....	(167)
第三节 CNKI数据库的使用简介.....	(171)
第四节 人大复印资料数据库使用简介.....	(174)
第五节 超星数字图书馆使用简介.....	(176)
第六节 外文数据库使用简介.....	(180)
第九章 特种文献及其检索	(196)
第一节 会议文献及其检索.....	(193)
第二节 学位论文及其检索.....	(203)
第三节 标准文献及其检索.....	(209)
第四节 专利文献及其检索.....	(212)
附录 1 美国化学文摘《Chemical Abstracts》摘录的网上期刊及网址	(227)
附录 2 美国《化学文摘》类目列表	(228)
附录 3 中国图书馆图书分类法(简表)	(231)
附录 4 在线工具书(On-Line Reference Books)	(237)
主要参考文献	(239)

第一章 信息检索的基本知识

第一节 信息的概念

随着科学技术的飞速发展,现代社会已进入信息时代。通信技术、计算机技术和声像技术等现代信息技术的发展及相互融合,拓宽了信息的传递和应用范围,使人们在广阔范围内可随时随地获取和交换信息。

那么,信息(Information)是什么呢?早在唐朝就有“梦断美人沉信息,目穿长路倚楼台”的诗句,这里的信息就是消息的意思,古朴的信息通常理解为信号和消息。随着科技的不断发展和人类的不断进步,信息的概念也有所深化。

迄今为止,信息还没有一个统一的定义,不同的作者站在不同的角度就有不同的理解。例如,《信息论》的创立者申农(C. E. Shannon)着眼于对信息的描述,把信息定义为“用于消除不定性的东西。”《控制论》的创立者维纳(N. Wiener)着眼于信息应用,对信息的定义是:“信息就是我们适合外部世界进行交流的内容的名称。”我们这里把信息定义为:信息是指应用文字、数据或信号等形式通过一定的传递和处理,来表现各种相互联系的客观事物在运动变化中所具有特征性的内容的总称。不同事物具有不同的存在方式和运动规律,从而构成了各种事物的不同的特征,发出不同的信息。总之,信息是对客观世界中各种事物的变化和特征的反映,是客观事物之间相互作用和联系的表征,是客观事物经过感知或认识后的再现。通俗地讲,信息就是通过信号带来的消息,信息普遍存在于自然界、社会界和思维界,信息与物质、能量共同构成了现代社会的三大资源,成为社会进步的强劲推动力。

第二节 文献的定义

文献就是用文字、图形、符号或声像、视频等技术手段记录知识信息的物质载体,或曰固化在物质载体上的知识信息。

文献的构成主要有三个要素:第一,有一定的知识内容;第二,有用以保存和记录知识的技术手段;第三,有记录知识的物质载体。这三个要素是文献的内涵。不同的知识内容、不同的记录方式和不同的物质载体,可形成不同类型的文献,这就是文献概念的外延。

文献是记录知识信息的物质形式,也是借以传递知识信息的工具。由于有文献的存在,人类的知识才得以保存和传播,人类的科学技术和文化才得到继承和发展。文献不仅包括书刊等印刷型出版物,还包括古代的甲骨文、竹简、帛书等,以及当今的声像出版物、电子出版物和Internet网上的信息。

第三节 文献的特点

一、文献数量大,增长速度快

随着科学技术和社会生活的飞速发展,人类知识总量在迅速猛增。作为存储、传播知识的载体文献,随着知识量的增加其数量也在激增,而且增长速度很快。据统计,目前全世界每年

出版各种文献总量约 12000 万册,平均每天出版文献约 32 万件。表 1-1 是美国《化学文摘》报道文献量的变化情况,从中不难看出文献数量的增长之快。

表 1-1 美国《化学文摘》报道文献量的变化情况

第一个百万条	1907—1946 年	历时 40 年	第五个百万条	1972—1975 年	历时 4 年
第二个百万条	1947—1960 年	历时 14 年	第六个百万条	1976—1978 年	历时 3 年
第三个百万条	1961—1967 年	历时 7 年	第七个百万条	1979—1980 年	历时 2 年
第四个百万条	1968—1971 年	历时 4 年	第八个百万条	1981—1982 年	历时 2 年

文献数量的激增,一方面表明文献信息资源的丰富,但同时也产生了“文献信息污染”,给人们选择、利用文献,获取所需信息造成了障碍。

二、文献分布集中又分散

随着现代科学技术的日益综合与细化,各学科之间的严格界限日趋淡化,学科之间的相互联系、交叉渗透逐渐增强。这使文献的分布呈现出集中又分散的现象,即某一专业的大部分文章发表在少量的专业性期刊中,而另外一部分文章则刊载在大量的相关专业,甚至不相关专业的杂志中。英国文献计量学家布拉德夫发现:如果将载有某一学科相关论文的期刊按载文量多少排列,并等分为几个区,这几个区的期刊数量分别为 $n_1, n_2, n_3, \dots$, 则有: $n_1 : n_2 : n_3 : \dots \approx 1 : \alpha : \alpha^2 : \dots$ (其中 α 是与学科有关的常数, $\alpha \approx 5$)。例如,利用英国《科学文摘》统计了“日本光纤通信”方面的论文 322 篇,发现共发表在 66 种杂志上,其中有 107 篇文章发表在 3 种杂志上(核心期刊),约有 33% 的文章发表在另 12 种期刊上(相关期刊),还有近 1/3 的文章则发表在另 51 种期刊上(边缘期刊)。文献的这种分布现象是普遍存在的,它提示人们在收集文献信息时,应首先选择本专业的核心期刊。

可以使用有关的工具来了解某专业的核心期刊,例如,由北京大学图书馆等单位合编的《中文核心期刊要目总览》(2004 年版)、《国外科学技术核心期刊总览》(2004 年版)、《国外人文社会科学核心期刊总览》(2004 年版)等。

三、文献时效性增强

随着科学技术的发展,新知识、新理论、新技术、新产品层出不穷,加速了知识的新陈代谢,随之造成了文献的新陈代谢,使文献老化加速。国外有人统计各类文献的平均寿命为:图书 10~20 年、科技报告 10 年、学位论文 5~7 年、期刊 3~5 年、标准文献 5 年、产品样本 3~5 年。

通常用文献的“半衰期”来描述文献老化情况。所谓文献的半衰期是指某学科领域目前尚在使用的全部文献中,较新的一半所出版的年限。国外有人统计不同学科文献的半衰期为:地理学 16.1 年、地质学 11.8 年、数学 10.5 年、植物学 10 年、化学 8.1 年、生理学 7.2 年、机械工程 5.2 年、社会科学 5 年、化工 4.8 年、物理学 4.6 年、冶金学 3.9 年、生物医学 3 年。由于各国科技发展水平不同,相应的文献寿命也不相同。

四、文献内容交叉重复

现代科学技术综合交叉、彼此渗透的特点,导致知识的产生和文献的内容也相互交叉、彼此重复。具体表现为:

(1) 同一内容的文献以不同文字发表。据统计,当前世界上每年翻译图书约占图书出版总量的 10%;一些重要的核心期刊被译成多种文字在不同国家出版;同一项发明可以向多个国家申请专利,使专利说明书的内容重复现象严重。据世界知识产权组织统计,世界各国每年

公布的专利说明书的重复率达 65%~70%。

(2) 同一内容的文献以不同形式出版。据调查,有 40% 的会议论文会后经过整理、修改发表在期刊上;有相当数量的科技报告后来发表在期刊上,如美国的 NASA 报告有 80%、AD 报告有 60%、美国科学基金会报告有 95%、美国农业部的科技报告有 80%,既出版单行本又发表在期刊上;随着新型载体文献的普及应用,许多文献既出版印刷版,又有缩微版、电子版等。如《纽约时报》、《泰晤士报》、《读卖新闻》等既有印刷版,又有缩微版;《中国大百科全书》、《不列颠百科全书》、《工程索引》、《科学文摘》、《科学引文索引》等既有印刷版,又有电子版。

(3) 在激烈的商业竞争中许多畅销书内容雷同,选题重复。此外,再版、改版文献数量不断增多,如 1998 年我国出版图书 130613 种,其中新版图书 74719 种,仅占 57%。

五、文献载体及语种增多

随着声、光、电、磁等技术和新材料的广泛应用,新型文献载体不断涌现。传统的纸张型文献已失去了一统天下的局面,多种载体的文献相互依存、相互补充、共同发展已成为趋势。新型的非纸张型文献如缩微型、机读型、视听型等,或增大了信息存储密度、延长其保存时间;或加快了信息传递与检索速度,实现了资源共享;或令人闻其声、见其形,使文献家族日臻繁荣。

各国文献所用语种不断增多。据 ISDS(International Serials Data System)报道,世界上连续出版物使用的语种多达 144 种,常用的就有 10 多种,其中英文约占 42%。

第四节 文献的特征

一篇文章、一本书、一份报告等一般都有以下特征。

一、外表特征

外表特征通常包括题目、作者、作者工作单位,专利和科技报告,还有专利号或报告号等。这些可以表征一篇特定文献的特征可以在文献的封面或扉页,即不打开书本,或不看文献的具体内容就可以确定一篇文献。

二、内容特征

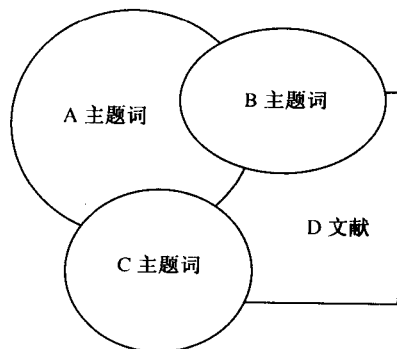
假如我们深入到文献内容中间,则可以发现还可用另外两种方法来表征它:

(1) 一般一篇文献都是论及某一方面的特定问题的,也就是说,与论题相关的词出现的频率较大。以前的研究表明,无论哪一种类型的文献,若对文献中出现的词进行频率统计的话,会发现所有的词可分为三类。

① 文献中出现频率最高的词是冠词、介词和连词等,即其本身没有具体含义的词,如 a、an、the、this、that、or、and、in、on、with 等;

② 绝大部分词在文献中出现的频率较低;

③ 在文献中出现的频率既不高也不低的词,在文献中约 3~20 个之间,这些词恰恰是与文献的主题相关度较大的词,我们称之为文献的主题词或关键词。主题词或关键词即称为文献的主题特征,其与文献的关系可用图 1-1 示出。



例如,有一篇文献是关于“用催化剂来合成聚乙

图 1-1 文献主题特征与文献的关系图

烯”，则我们可选出三个词：催化剂、聚乙烯、合成，通过它们之间的组合就可大致表示出这篇文献的主要内容了。

由于作者对自己的文献较为熟悉，因而在具体操作时关键词均由作者本人给出。

(2) 另外，一篇文献还可以按照分类方法进行归类，如《中图法》(详见附录 3)：

N	O	P	Q	R	
自然科学总论	数、理、化	天文、地理	生物	医、卫	
O1	O3	O4	O6	O7	
数学	力学	物理	化学	晶体	
O11	O12	O13			
古典数学	初等数学	高等数学			
O121	O122	O123	O124		
算术	初等代数	初等几何	三角		

这样，我们可以把某一篇文章归入某一类中，如有关三角方面的文献可归入 O124 中，O124 即可称为该文献的分类特征。

综上所述，文献内容特征分为两个方面：

① 主题特征：主题词、关键词等；

② 文献分类特征：分类号、模糊对应等。

文献的外表特征与文献是一一对应的，即一组外表特征只对应一篇唯一的文献，而文献的内容特征与文献却是一种模糊的对应关系，即一篇文献有多个主题词(关键词)或分类号，一个主题词(关键词)或分类号也可对应多篇甚至几百篇文献。正因为存在这两种对应关系，使得两种特征的用途大不相同。利用外表特征只能检出很少的文献，有时只用于特定情况下(如已经知道作者名等)。利用内容特征一次能检出一批文献，这对研究者来说是极其有用的。

第五节 文献的级别、种类和载体形式

一、文献的级别

一次文献(primary document)：是指作者以本人的研究成果为基本素材而创作或撰写的文献，不管创作时是否参考或引用了他人的著作，也不管该文献以何种物质形式出现，均属一次文献。大部分期刊上发表的文章和在科技会议上发表的论文均属一次文献。

二次文献(secondary document)：是指文献工作者对一次文献进行加工、提炼和压缩之后所得到的产物，是为了便于管理和利用一次文献而编辑、出版和累积起来的工具性文献。检索工具书和网上搜索引擎是典型的二次文献。

三次文献(tertiary document)：是指对有关的一次文献和二次文献进行广泛深入的分析研究综合概括而成的产物。如大百科全书、辞典等。

也有研究者在以上分类基础上再加上零次文献，它是指未经过任何加工的原始文献，如实验记录、手稿、原始录音、原始录像、谈话记录等。零次文献在原始文献的保存、原始数据的核对、原始构思的核定(权利人)等方面有着重要的作用。

二、文献的种类

我们主要讲的是各级文献，它是由各种形式、性质、特点的出版物组成的，大致可归纳为 10 大类。现分述如下：

1. 图书

图书是一种传统的、成熟定型的出版物,如专著、丛书、教科书、论文集、工具书等等。每种图书都有一个主题,反映的知识内容力求完整、系统和成熟,是编著者长期研究的成果和学识的积累,是传播知识、培养人才的主要读物。若要对某学科或某专题获得较全面、系统的知识,或对不熟悉的问题要获得基本的了解时,选择图书是行之有效的方法。图书的特点是:全面、系统,理论性强,技术成熟可靠,但是出版周期长,反映的知识内容相对滞后,一些新理论、新观点、新技术、新方法等不能及时反映。图书按功用性质可以分为:阅读性图书和工具书两大类。阅读性图书包括教科书(Textbook)、专著(Monograph)、文集(Anthology)等。它提供系统、完整的知识,有助于全面、系统地了解某一领域的历史发展与现状,将人们正确地引入自己所不熟悉的领域。工具书(Reference book)包括词典(Dictionary)、百科全书(Encyclopedia)、手册(Handbook)、年鉴(Yearbook)等。它提供经过验证、浓缩的知识,是信息检索的工具。图书在各种类型的图书馆有广泛的收藏。

图书在各种论文末的参考文献或题录性检索工具中通常著录成如下例所示的格式:

W. V. Etten, Fundamentals of optical fiber communication, London: Prentice—Hall (1991), 407pp

图书的著录特点是:有书名,有著者,有的还有编者;还有出版地、出版社名和出版年份;非第一版的图书有版次;有的图书还给出国际标准书号(ISBN)。这四点中“出版项”较重要。

识别图书的主要依据有:书名、著者、出版地、出版社、出版时间、总页数、国际标准书号(ISBN)等。其中 ISBN 由 10 位数字分成 4 段组成,各段依次是:地区或语种号-出版商代号-书名号-校验码。其中地区或语种号:英、美、加、南非等英语区为 0,其他英语区为 1,法语区为 2,德语区为 3,日本是 4,俄语区为 5,中国为 7,印度等为 8,东南亚地区为 9。如:7-302-02372-7,表示中国代号为 302 的(清华大学)出版社,出版的一种图书,其书号为 02372,该书的校验码为 7。

2. 期刊

期刊也称杂志,是指那些定期或不定期连续出版的文献载体,它是由专门的编辑机构编辑出版的一种连续出版物(Serials)。期刊能及时反映学科新观点、科研新成果、社会新动态,是科研人员了解科研动态、前沿信息和进展情况的首选文献。

期刊的特点是内容新颖、信息量大、出版周期短、传递信息快、传播面广、时效性强,能及时反映国内外各学科领域的发展动态。据统计,科技人员所获取信息的 65% 以上来源于期刊,它是十分重要和主要的信息源和检索对象。

期刊一般有同样的名称,按一定时间编定卷号,每一卷又分若干期。其类型有:①学术性、技术性刊物。如 Acta(学报)、Journals(杂志)、Annals(纪事)、Bulletin(通报)、Transactions(汇刊)、Proceedings(会刊)、Review(评论)、Progress(进展)等。②快报型刊物。如 Communication(通讯)、Bulletin(公报)等。③消息性 News(新闻)等。④资料性 Data(数据)、Event(事件)等。

期刊论文的著录格式如下例所示:

H. Tohyama, A plasma image bar for an electrophoto—graphic printer, Journal of the Imaging Science, Vol. 35 No. 5, 330—3(Sept-Oct 1991)

期刊论文著录的特点是:有作者,有时有篇名;期刊名称常常缩写,有的还以斜体给出;必定有卷号,有的有期号。这三点中“卷期号”较为重要。

识别期刊的主要依据有:期刊名称,期刊出版的年、卷、期,国际标准刊号(ISSN)等。其中 ISSN 由 8 位数字分两段组成,如 1000-0135,前 7 位是期刊代号,末位是校验号。我国正式出版的期刊都有国内统一刊号(CN),它由地区号、报刊登记号和《中图法》分类号组成,如 CN11-2257/G3。地区号依《中华人民共和国行政区划编码表 GB2260-82》取前两位,如北京 11、天津 12、上海 31、辽宁 21、吉林 22 等。

3. 科技报告

科技报告是描述一项研究进展或取得的成果,或一项技术研制试验和评价结果的一种文体。科技报告所反映的科研和技术革新成果比期刊论文快。由于它是科研进展和试验过程的真实纪录,往往能反映一个国家在某一学科的科技水平,故能起到直接的借鉴作用。可以说,报告是一种典型的机关团体出版物,常用的报告可分为:

按报告的形式:Report(报告书),Technical Notes(技术札记),Memorandum(备忘录),Papers(论文),Bulletin(通报),Technical Translations(技术译丛),Special Publications(特种出版物)。

按研究的进展情况:Primary Report(初步报告),Progress Report(进展报告),Interim Report(中间报告),Final Report(最终报告)。

按报告的保密程度:Top Secret(绝密),Secret(秘密),Confidential(保密),Restricted(限制发行),Unclassified(公开),Declassified(解密)。

科技报告的特点是内容新颖、详细、专业性强、出版及时、传递信息快,每份报告自成一册,有专门的编号(即报告号,通常由报告单位缩写代码+流水号+年代号构成),发行范围控制严格,不易获取原文。因科技报告反映新的研究成果,故它是一种重要的信息源,尤其在某些发展迅速、竞争激烈的高科技领域,人们对其需求更为迫切。在我国,国家图书馆、上海图书馆、中国科技信息研究所和国防科技信息研究所等收藏有较全面的科技报告。

科技报告常见的著录格式如下例所示:

J. S. Bymes, "Application of approximation theory in antenna design, signal processing and filtering", Final report. AD-A244,725 Prometheus, Inc., Sharon, MA. 1991

科技报告著录的特点是:有表示报告的词,如 Report, Memorandum 等;有报告号,如本例中的 AD-A244725。这两点中“报告号”较重要。

识别科技报告的主要依据有:报告名称、报告号、研究机构、完成时间等。

目前全世界每年发表科技报告数量庞大,其中绝大多数产自发达国家,较著名的有美国政府的四大报告(PB、AD、NASA、DOE)、英国航空委员会 ARC 报告、法国原子能委员会 CEA 报告、德国航空研究所 DVR 报告等。我国从 1963 年开始科研成果的正式报道工作。

4. 会议文献

会议文献是指学术会议文献,它往往反映出科学技术的发展趋势,其特点是与最新成果的间隔时间短,但其内容与期刊相比可能不太成熟。会议类型大致可分为国际会议、全国性会议及地区会议等。会议和会议文献常用的主要名称有 Conference(大会)、Meeting(小会)、Symposium(讨论会)、Proceedings(会议录)、Paper(单篇论文)、Transactions(汇刊)等。

会议文献按出版时间主要可分为:会前文献和会后文献。会前文献主要有会议论文预印本和会议论文摘要。会后文献是会后经整理出版的文献,如会议录(Proceedings)、会议论文集(Symposium)、会议论文汇编(Transactions)、会议丛刊、丛书等。目前,有许多学术会议在互联网上举行。由于会议文献的出版形式多样,会议文献的入藏分散,检索及获取不如图书、

期刊那样容易。通常,以图书或期刊形式公开出版发行的会议文献,在大型图书馆和省级以上信息研究所收藏较多。

会议文献常见的著录形式如下例所示:

T. Sellis: "Performance of DBMS implementation of production systems", Proceedings of the 2nd International IEEE Conference on Tools for Artificial Intelligence Herndon, VA USA, 6-9 Nov. 1990, (Alamitons, CA, USA: IEEE Computer Society Press 1990), 393-9

会议文献著录的特点是:有表示会议的专门用词,如 Conference, Symposium, Convention, Workshop, Meeting, Congress, Assembly 等;有表示会议录的一些词,如 Proceedings of ..., Collection of ...;有的有会议召开的地点、届次、时间,以及会议录的出版社、出版地、出版时间等。这三点中“会议的专门用词”最为重要。

识别会议文献的主要依据有:会议名称、会址、会期、主办单位、会议录的出版单位等。

5. 专利文献

专利文献在实行专利制度的国家,凡是本国或外国的个人和企业有了创造发明,都可以根据专利法的规定,向本国或外国专利局提出申请,经审查合格,批准授予在一定年限内享有创造发明成果的权利,并在法律上受到保护,这样一种受到法律保护的技术专有权利叫做专利。专利是与产业活动密切相关的,因而其实用性非常强。

根据专利的技术水平和应用情况,其类型有:invention patent(发明专利)、utility patent(实用专利)和 design patent(设计专利)等。

专利文献主要由专利局的专利文献馆和省级以上科技信息所收藏。专利文献的主要内容是专利说明书。

专利说明书常见的著录形式如下例所示:

B. D. Dayton, "Differential amplifier apparatus", US Patent No. 5095282, 1992

专利说明书著录的特点是:通常有表示专利的词(Patent);有专利号。专利号按国际规定由两个字母表示的国家名称和其后的顺序号构成。这两点中“专利号”较重要。

6. 标准文献

标准化主要包括三个方面:

(1) 产品标准化,即产品质量要符合技术规定。

(2) 产品规格化和系列化,即产品按型号大小来分档,成系列的发展,从而保证以较少的品种来满足广泛的要求。

(3) 零部件通用化,即同类机型零部件,特别是易损件要做到最大限度的通用互换性。

经过公认权威当局批准的标准化工作成果,可以采用文件形式或规定基本单位(物理常数)这两种形式固定下来,以文件形式出现的标准化工作成果,就是标准文献。

标准文献都有标准号,它通常由国别(组织)代码+顺序号+年代组成,如 ISO3297-1986。我国的国家标准分为强制性的国标(GB)和推荐性的国标(GB/T),如 GB18187-2000、GB/T2662-1999;行业标准代码以主管部门名称的汉语拼音声母表示,如 JT 表示交通行业标准;企业标准编号:Q/省、市简称+企业名代码+年份。

7. 学位论文

学位论文是指高等学校或研究机构的学生为取得某种学位,在导师的指导下撰写并提交的学术论文,它是伴随着学位制度的实施而产生的。学位论文有博士论文、硕士论文、学士论文之分,其研究水平差异较大,博士论文论述详细、系统、专深,研究水平较高,参考价值大。

识别学位论文的主要依据有:学位名称、导师姓名、学位授予机构等。

8. 产品资料

产品资料是对已经投入生产的产品作介绍的资料。产品资料包括产品样本、产品目录、产品说明书、厂商介绍、厂刊、技术座谈资料等。好的产品说明书含有丰富的内容,包括产品规格、产品特点、产品专利号等多种对生产有用的信息。

9. 科技档案

科技档案是研究部门和生产单位在科学研究、生产实践中所形成的有具体工程对象或科研对象的真实记录材料,有图样、照片、图表、原始记录及其复制本文献。科技档案包括任务书、协议书、审批文件、研制计划、方案、技术调查材料、技术措施、试验项目、试验方案、试验数据、设计图纸、工艺记录等等。科技档案记录内容真实、准确、可靠,它反映了生产和科技活动的全过程及最后结果。科技档案是生产科研中能借以积累经验、吸取教训和提高质量的重要参考文献。

10. 报纸

报纸是指有固定名称、刊期、开本,以新闻报道为主要内容,每周至少出版一期的散页连续出版物。它对于了解新闻和学科前沿及水平很有益处。

在十大文献中,最重要的是图书、期刊、会议论文和专利。

三、文献的载体形式

1. 印刷型

印刷型文献是以纸张为载体的出版物,是传统的记录知识、信息的方式。其优点是便于阅读,可广泛流传;缺点是体积大,占用空间多,难以实现机械化、自动化。

2. 缩微型

缩微型文献是以感光材料为载体,利用摄影技术将文献影像体积缩小记录在胶卷或胶片上。缩微型文献可将很多文献集中在小小胶片上储藏,其特点是:体积小、容量大、成本低,便于复制、携带、保存,节省库位。但缩微文献在阅读上不太方便,使用时须借助阅读机,不像印刷型文献随时可看。

3. 视听型

视听型文献包括:唱片、录音带、录像带、科技电影、幻灯片。如心脏杂音可录入唱片,将手术过程拍成影片,能听其声、观其形是视听文献的优点,便于理解、掌握。视听文献容易保存,可以反复使用。随着我国社会的不断进步,利用这类文献的人会越来越多。但视听文献也必须借助于录音机、录放机、放映机、幻灯机、计算机等来阅读。

4. 机读型

机读型是用计算机阅读的文献,通过各种编码并根据一定程序把文献录在计算机的磁带、磁盘、光盘、硬盘等载体上。

思考题:

1. 文献定义及其特征。
2. 文献的种类有几种? 如何识别图书、期刊、会议论文、专利文献、科技报告?

第二章 信息检索原理

第一节 信息检索原理、类型和方法

一、信息检索的原理

信息存储工作包括文献收集、信息加工整理,最后形成有序的、具有存储标志的文献信息数据库;信息检索工作则包括用户提出课题检索要求,通过分析形成检索标志,然后到文献信息数据库中进行检索标志与存储标志的匹配,凡是匹配成功的,相应信息款目即为中选检索结果。

因而,所谓信息检索原理,简单地讲,就是检索提问标志与存储在检索工具中的文献特征标志进行比较或称匹配,然后提取相符合的文献信息的检索方法。在机检中所采用的检索原理通常是依靠逻辑运算的布尔检索方式,这种方法是运用以“与”(and)、“或”(or)以及“非”(not)为主的逻辑运算来检索所需信息的。

二、信息检索类型及方式

信息检索,可根据检索时所使用不同的工具和手段,划分为手工检索和计算机检索两大类型。

1. 手工检索

手工检索多以书本式或卡片式检索工具为主,手工检索需要了解标引规则,按规则进行各项的著录方式,检索者根据文献标引规则查阅有关文献,是计算机检索方法的基本功之一。手工检索能了解各类检索刊的收录范围、专业覆盖面、特点和编制要点,可以提高查全率和查准率。因此,手工检索仍不失为较好的检索手段,专题的检索工具刊可与综合检索工具刊(或数据库)相互补充使用,以免漏检,影响查全率和查准率。手工检索也便于检索策略的制订和修改,手工检索过程发现问题,可以及时修改和提出。合乎逻辑提问式,选准检索词,再利用计算机检索,缩短机检时间,查全查准。利用手工检索,灵活性高,费用低,又能与机检互为补充,在今后相当长时间内,手工检索方法仍是重要的检索手段。

2. 计算机信息检索

计算机信息检索就是指人们在计算机检索网络或终端上,使用特定的检索指令、检索词和检索策略,从计算机检索系统的数据库中检索出所需要的信息,然后再由终端设备显示和打印的过程。为实现这种信息检索,必须事先将大量的原始信息加工处理,存储在各种信息载体上待用,所以计算机信息检索广义上讲包括信息的存储和检索两个方面。

信息存储的过程就是将所选中的一次文献进行主题分析、标引和著录,按一定格式输入计算机,构成可供机读数据库记录及文献特征标识,这相当于编制手工检索用的文摘索引等检索工具。

信息检索则是查询的过程。用户对检索课题加以分析,明确检索范围,弄清主题概念,然后用系统语言来表示主题概念,形成检索标识及检索策略,输入到计算机进行查找。这一查找的过程实际上是计算机自动比较匹配的过程,当检索标识、检索策略与数据库中的信息特征标

志及其逻辑组配关系相一致时,则属“检索命中”,即找到了符合要求的信息。检索结果可以联机或脱机打印输出。

计算机检索不仅能够跨越时空,在短时间内查阅大型数据库,还能快速地对几十年前的文献资料进行回溯检索,而且大多数联机或网络检索系统数据库中的数据更新速度非常快,信息用户通过计算机信息检索随时可以得到更新的信息。

目前科技发展迅速,知识和信息产量呈指数级增长,国际间的合作和联系与各国科技发展密切相关,学科间相互交叉、相互渗透,边缘学科发展很快,科学技术的综合性越来越强,因此培养 21 世纪人才,必须借助于现代信息技术。Internet 是世界范围的联网信息系统,通过此国际计算机互联网络,能及时获取所需信息,与世界交流合作,及时进入相关学科领域,参与世界的科技竞争。因此,兴建信息高速公路,开发网络信息技术,加快计算机联网,利用全球网络信息,开发国内信息资源极为重要。

3. 信息检索方式

人们获取信息的方式通常有两种:直接检索、间接检索,本文所描述的检索原理主要针对后者而言。

(1) 直接检索

直接检索就是从浏览、阅读文献中(主要指一次文献)获取所需要的情报信息过程,它是研究者习惯使用的方法,具有以下几个突出优点:

- ① 易于把握文献信息实质内容;
- ② 可直接对信息内容进行判断取舍;
- ③ 简便易行;
- ④ 信息获取速度快,时效高。

但直接检索信息方式有难以避免的严重缺陷:

- ① 信息传递与交流的速度较慢;
- ② 所获信息量只占信息总量的很小部分;
- ③ 检索者受语言、地理等影响较大。

(2) 间接检索

间接检索就是以检索工具(主要是二次文献和部分三次文献)为导向,进而查找原始文献以获得所需信息的过程。其优越性在于:

① 使盲目的分散检索成为有目的的集中检索。因为检索工具将分散在不同学科、不同类型、不同语种,但主题内容相同的文献集中在一起,这样就可避免直接检索的分散性、盲目性,大幅度提高检索效率;

② 检索工具中的信息源不限于某一个或几个信息机构,这样就可以为读者提供广泛的信息来源;

③ 提供有规律的检索途径,因而检索者只要掌握检索工具所提供检索途径的规律性,就可以实现有效的信息检索;

④ 消除语言障碍,因为检索工具通常只用一种文字(主要是英文)写成,检索者只需掌握常用文种,即可实现多语种信息检索;

⑤ 检索工具通常提供相关主题的题录或文摘而不是全文,因而可加快检索速度。

由此可见,间接检索相对直接检索更科学,在实践中若能有机结合,则能相得益彰。

第二节 信息检索工具

一、检索工具的涵义

检索工具是指用以报道、存储和查找文献线索的工具。它是附有检索标识的某一范围文献条目的集合,是二次文献。

一般说来,检索工具应具备以下五个条件:

- (1) 明确的收录范围;
- (2) 有完整明了的文献特征标识;
- (3) 每条文献条目中必须包含有多个有检索意义的文献特征标识,并标明供检索用的标识;
- (4) 全部条目科学地、按照一定规则组织成为一个有机整体;
- (5) 有索引部分,提供多种必要的检索途径。

二、检索工具的类型

目前可供人们使用的检索工具有很多,不同的检索工具各有特点,可以满足不同的信息检索的需求。

检索工具有不同的分类方法,按加工文献和处理信息的手段不同可分为:手工检索工具和机械检索工具。按照载体形式不同可分为:书本式检索工具,磁带式检索工具,卡片式、缩微式、胶卷式检索工具。按照著录格式的不同可将检索工具分为以下四种类型。

1. 目录型检索工具

目录型检索工具是记录具体出版单位、收藏单位及其他外表特征的工具。它以一个完整的出版或收藏单位为著录单元,一般著录文献的名称、著者、文献出处等。目录的种类很多,对于文献检索来说,国家书目、联合目录、馆藏目录等尤为重要。

2. 题录型检索工具

题录型检索工具是以单篇文献为基本著录单位来描述文献外表特征(如文献题名、著者姓名、文献出处等),无内容摘要,是快速报道文献信息的一类检索工具。它与目录的主要区别是著录的对象不同。目录著录的对象是单位出版物,题录的著录对象是单篇文献。

3. 文摘型检索工具

文摘型检索工具是将大量分散的文献,选择重要的部分,以简练的形式做成摘要,并按一定的方法组织排列起来的检索工具。按照文摘的编写人,可分为著者文摘和非著者文摘。著者文摘是指按原文著者编写的文摘;而非著者文摘是指由专门的熟悉本专业的文摘人员编写而成。就其摘要的详简程度,可分为指示性文摘和报道性文摘两种。指示性文摘以最简短的语言写明文献题目、内容范围、研究目的和出处,实际上是题目的补充说明,一般在100字左右;报道性文摘以揭示原文论述的主题实质为宗旨,基本上反映了原文的内容,讨论的范围和目的,采取的研究手段和方法,所得的结果或结论,同时也包括有关数据、公式,一般500字左右,重要文章可多达千字。

4. 索引型检索工具

索引型检索工具是根据一定的需要,把特定范围内的某些重要文献中的有关款目或知识单元,如书名、刊名、人名、地名、语词等,按照一定的方法编排,并指明出处,为用户提供文献线索的一种检索工具。

索引的类型是多种多样的,在检索工具中,常用的索引类型有:分类索引、主题索引、关键