

滕胜娟 唐丽菲 / 主编

现代 信息 检索



 中国纺织出版社

现代信息检索

滕胜娟 唐丽菲 主编



中国纺织出版社

内 容 提 要

本书为《图书馆·文献检索》的姊妹篇。在全球信息化的今天,如何以最快的速度查找和利用现代信息,已成为提高市场经济环境下生存、竞争和发展能力的关键。本书主要介绍光盘数据库检索、网络信息检索、国际联机检索等计算机信息检索的基本原理和方法,几种常用的手工检索工具。机检与手检相结合,突出检索技能的训练与培养;取材新颖精炼,注重实用。可用作普通高等院校文献检索课程教材,又可作为工程技术人员和科研人员索取信息和读书治学的参考书籍。

图书在版编目(CIP)数据

现代信息检索/滕胜娟,唐丽菲主编. —北京:中国纺织出版社,2002.1

ISBN 7-5064-2156-9/G·0075

I. 现… II. ①滕…②唐… III. 情报检索 IV. G252.7

中国版本图书馆 CIP 数据核字(2001)第 080041 号

策划编辑:卞 葆	曹炳鎔
责任编辑:孙 玲	责任校对:余静雯
责任设计:胡雪萍	责任印制:初全贵

中国纺织出版社出版发行

地址:北京东直门南大街 6 号

邮政编码:100027 电话:010—64168226

<http://www.c-textilep.com>

E-mail: faxing@c-textilep.com

中国纺织出版社印刷厂印刷 各地新华书店经销

2002 年 1 月第一版第一次印刷

开本:850×1168 1/32 印张:6.5

字数:132 千字 印数:1—5000 定价:15.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

前 言

在全球信息化的今天,电子出版物正挑战传统出版物,计算机检索正挑战传统手工检索,它对传统的信息组织、检索和获取方法构成了很大冲击。信息处理智能化、网络通信普及化和多媒体技术实用化已是社会发展的必然趋势,使世界范围内的信息交流、资源共享成为可能,从而大大拓展了人类的信息空间。为培养读者查找与利用现代信息的能力,提高在现代信息环境下的生存和发展水平,基于近年来的学习、研究与教学实践,我们编写了《现代信息检索》。

本书特点在于以论述国内光盘数据库为主,兼顾国外光盘数据库;以论述计算机信息检索为主,兼顾手工检索工具,将机检与手检融为一体。该书不仅可作为高等院校文献检索课程的教材,而且也可作为广大读者自学检索方法的指南。

本书执笔分工如下(按章节顺序):

滕胜娟	第 1 章
刘 诚	第 2 章(2.1~2.6)
唐丽菲	第 2 章(2.7~2.12)
唐丽菲	第 3 章

兰 曦 第 4 章(4.1~4.3)

唐丽菲 第 4 章(4.4)

滕胜娟 第 5 章

全书由滕胜娟、唐丽菲统稿、定稿,高云主审。

由于我们的编写水平有限,书中如有疏漏和不妥之处,恳
请广大读者和用户指正。

编 者

2001.10.

目 录

第1章 计算机信息检索概述	(1)
1.1 信息检索	(1)
1.2 计算机信息检索及其意义	(7)
1.3 计算机信息检索语言	(14)
1.4 计算机信息检索系统	(22)
1.5 计算机信息检索的原理和方法	(29)
第2章 光盘数据库检索	(38)
2.1 概况	(38)
2.2 中国学术期刊光盘数据库	(43)
2.3 中文期刊数据库(题录文摘版)	(53)
2.4 中国纺织文摘数据库	(58)
2.5 全国报刊索引数据库	(61)
2.6 中国人民大学复印报刊资料全文数据库	(66)
2.7 外文期刊光盘数据库(题录文摘版)	(70)
2.8 化学文摘光盘数据库 CA(Chemical Abstract)	(72)
2.9 工程索引光盘数据库 EI(The Engineering Index)	(75)
2.10 科学文摘光盘数据库 INSPEC(Information Service in Physics,Electro - Technology, Computer and Control)	(79)
2.11 科学引文索引光盘数据库 SCI(Science	

	Citation Index)	(82)
2.12	博、硕士论文光盘数据库 PQDD(ProQuest Digital Dissertation)	(85)
第3章	网络信息检索	(88)
3.1	因特网概述	(88)
3.2	网络信息检索	(101)
3.3	网上专题信息检索	(122)
第4章	国际联机检索	(148)
4.1	国内外主要联机检索系统	(148)
4.2	检索算符及检索方法	(153)
4.3	DIALOG 系统检索	(156)
4.4	OCLC 系统检索	(158)
第5章	手工检索工具	(162)
5.1	《纺织文摘》	(162)
5.2	英国《世界纺织文摘》	(166)
5.3	美国《纺织技术文摘》	(170)
5.4	英国《世界专利索引》	(175)
5.5	美国《化学文摘》	(185)
5.6	美国《世界会议》	(195)
参考文献		(199)

第 1 章 计算机信息检索概述

1.1 信息检索

1.1.1 信息检索的定义

信息检索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。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,相当于信息查寻。

根据检索对象的不同,信息检索可分为三种:数据检索、事实检索、文献检索。

数据检索:检索结果是数据,例如某种金属的熔点,某种材料的电阻。另外,有关计算公式、数据图表、化学分子式等都属于数据检索范畴。

事实检索:检索结果是事实结论。凡查询某一事物(事件)的性质、定义、原理以及发生的时间、地点、过程等,都属于事实检索的范畴。诸如某类产品由哪些厂家生产,哪个牌号最好等都属于事实检索。

文献检索:检索结果是文献资料。凡是查找某一课题、某一著者、某一地域、某一机构、某一事物的有关文献的出处和收藏单位等,均属于文献检索的范畴。例如要系统地收集有关“纺织印染厂

对废水处理的计算机控制系统”的文献,即属于文献检索。

在三种信息检索类型中,文献检索是最主要、最基本的形式,是信息检索中最重要的部分。

文献检索可分为手工检索(简称手检)和计算机检索(简称机检),手检是基础,机检是发展方向。

1.1.2 文献的类型与识别

文献是人类进行思想交流、文化交流的一种最重要的信息。“文献是记录有知识的一切载体”。构成文献有两个最主要的要素:其一,要有一定的知识内容,没有记录任何知识内容的纸张、录音带等不能称为文献;其二,要有用以记录知识的物质载体。从这个意义上讲,存在于人们头脑中的知识不能称为文献,只有将知识用文字、图像、数码等各种符号,采用书写、印刷或其他诸如光学、电磁学等方法记录在一定物质载体上,才叫文献。

文献的类型众多。从文献的定义不难看出,文献概念的外延非常广泛,可以按不同标准对它进行分类。

1.1.2.1 按文献的出版形式划分

1. 手写型 指手书未经过付印的书信、笔记、会议记录、手稿等原始资料。这类文献具有重要的馆藏价值。

2. 印刷型 包括铅印、油印、石印、胶印。主要以纸张为载体,它的优点是便于阅读,便于流传;缺点是占据空间大,整理和保存起来比较困难。

3. 缩微型文献 是以感光材料为载体,以照相复制方式,将纸张文献缩微复制在感光材料上而成的文献,包括缩微胶卷和缩

微胶片。具有体积小、重量轻、存储密度高的优点,但要借助缩微阅读器才能阅读,不大方便,也不便于携带。

4. 视听型文献 又称声像型文献,是指以磁性材料或感光材料为载体,采用录音、录像或摄影技术直接记录声音信息或图像信息而形成的文献,包括唱片、录音带、录像带、幻灯片、电影片等。其特点是直观、真切、传递速度快、存储密度高。

5. 电子型文献 包括磁带、磁盘、光盘文献。它主要是通过编码技术、光磁电技术和计算机技术,把文字、图像、声音转换成数码,存储在磁带、磁盘、光盘上,“阅读”时,由计算机将数码还原成文字、图像、声音显示在屏幕上。电子型文献数据库是计算机信息检索的主要对象,其优点是体积小、检索方便、直观生动,便于远距离传输;缺点是需借助计算机阅读,不大方便,费用较高。

1.1.2.2 按文献信息的加工程度划分

1. 零次文献 是未经过付印的原始记录,包括书信、手稿、笔记、会议录等。零次文献具有重要的信息价值。

2. 一次文献 是根据科研成果撰写的首次发表的原始文献,由零次文献转化而来,如期刊论文、研究报告、专利说明书、会议文献等。

3. 二次文献 是将分散的无组织的一次文献经过加工整理、简化和排序等程序,形成系统的便于查找和利用的文献,即检索工具:书目、索引、文摘、辞典、百科全书等,它的主要作用是检索一次文献。

4. 三次文献 是指在利用二次文献的情况下,选用一次文献

内容而编写出来的成果,如文献指南、书目之书目、综述、述评等。

从一次文献到二次、三次文献,是一个由分散到集中,由无组织到系统化的过程。从信息检索的角度来说,一次文献是检索的对象,二次文献是检索的工具。

1.1.2.3 按文献出版特点划分

1. 图书 图书是由出版社和其他单位、个人出版的围绕某一个题目进行全面论述的出版物,它可以是原始文献,也可以是二次或三次文献,也可以是它们的混合体。其内容一般更系统全面,其结论更成熟定型,适合于学习型读者。

图书在检索工具书或者数据库中著录格式一般为:书名、作者、出版社名称、出版地、出版时间、ISBN(国际标准书号)。

2. 期刊论文 期刊是围绕某个专题定期或不定期连续出版的出版物。与图书相比,期刊论文的内容不全面系统,不成熟,论题很窄,但它的出版周期短,载文速度快,内容新颖、专深,适合于研究型读者。期刊论文是所有类型文献中数量最多的一种,占整个信息源的65%,大多数文摘索引类的检索工具都是以期刊论文作为摘录和检索对象。

期刊论文在检索工具书或者数据库中著录格式一般为:期刊论文题名、作者、期刊名缩写、卷、期、年、页或标有ISSN国际标准期刊号。

3. 科技报告 是研究单位或个人向上级或委托单位撰写的关于某个课题研究成果的正式报告。它不对外发行,也不连续出版,其内容新颖,选题尖端适用。科技报告的保密性较强,但到了

一定的时间,保密文献陆续解密。因此科技报告也是我国科技人员跟踪国外科技研究方向、现状以及发展趋势的重要途径之一。

科技报告在检索工具书或者数据库中著录格式一般为:科技报告号、报告入藏号。

4. 会议文献 学术与专业会议都是围绕某一学科或专业领域的新成就和新课题来进行交流、探讨,它是科研人员进行学术交流的重要场合。会议文献是指在学术与专业会议上宣读的论文和报告。因此,它代表了一门学科或专业领域最新的科研成果,反映着世界上科学技术发展的水平和趋势。

会议文献在检索工具或数据库中著录格式一般为:会议名称、开会时间、开会地点、出版时间等。判断是否为会议文献,还可根据表示会议特征的英文名称来决定。例如:Conference, Congress, Proceedings, Symposium, Paper 等。

5. 专利文献 指专利申请书、专利说明书、专利公报等与专利有关的一切文献。内容一般是新颖、先进、实用的技术。

专利文献在检索工具或数据库中著录格式一般为:专利名称、专利权人名称、专利国代号、专利号、优先权项、国际专利分类号、专利申请日期、申请号等。

6. 技术标准 技术标准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。它作为权威部门制定的规章性文献,有一定的法律约束力,其详尽性、完整性和可靠性,非其他类型文献可比,新陈代谢非常频繁。

技术标准在检索工具或数据库中,著录的显著特征是:技术标

准号。例如:GB4580—84(中国标准)。

7. 学位论文 学位论文是大学本科生、硕士生和博士生为取得学位而撰写的论文。由于它一般不对外发行,印数少,所以检索和获取原文比较困难。学位论文的质量参差不齐,其中的硕士论文、博士论文费时较长,问题较专,研究较深,对研究工作有一定参考价值。

学位论文在检索工具或数据库中,著录格式一般为:学位名称、颁发学位的大学名称及其地址、授予学位的年份等。例如:Phy.D.Dissertation,Stanford University,Stanford,CA1989。

8. 产品样本和产品目录 这是对定型产品的性能、构造原理、用途、使用方法和操作规程、产品规格等所作的具体说明,包括单项产品的样品(产品说明书)、企业产品一览、企业介绍、单项产品样本汇编、同行业产品一览表等。由于产品的不断更新,产品样本也容易过时。

9. 技术档案 技术档案是记述和反映一个单位或部门在基本建设、生产技术和自然科学研究等活动中,具有保存价值,并且按照归档制度,作为真实历史记录集中保管的技术文件材料。例如规划、计划、方案、任务书、协议书、技术指标、审批文件、研究计划、实验项目、设计图纸、工艺规程、技术鉴定、经济核算、发明记录数据等。

10. 政府出版物 这是各国政府部门及其所设立的专门机构发布出版的文件,内容为行政文件,如决议、法令、白皮书、统计资料等。各国一般都有专门机构(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。

11. 报纸 报纸是一种报道及时、内容范围广、文字通俗的文献信息。由于报纸积累信息量大,且杂乱无序,不借助检索工具查找起来很困难,因此,重要报纸都编有月度年度索引,有的报纸文章与期刊论文一起被摘录编排,形成报刊索引。

还有其他一些类型的文献,如手稿、文书、文稿、统计资料、地图等,这里不一一列举。

文献类型尽管多样,但几乎所有重要类型的文献有专门的收藏部门,有专门的检索工具,大部分被做成专门的数据库。例如美国 DIALOG 联机检索系统已含数百个大型数据库,包含的文献类型很多,各学科、各类型的文献都可以在其中找到。

1.2 计算机信息检索及其意义

计算机信息检索就是用计算机查找信息资料,它产生的背景是信息资料的不断增长和人们检索要求的不断提高,尤其是计算机软件技术的不断发展。

从 1920~1930 年代开始,特别是第二次世界大战后,科技成果以空前的速度增长。据估计,近 20 年科技成果的总和超过这之前人类历史 2000 年科技成果的总和,形成“知识爆炸”和“信息爆炸”。这主要表现在:第一,科技文献的数量大,增长快。据“世界科学情报系统”估计,全世界发表的科技论文,其增长速度在 1950 年代以前大约每 15 年增长一倍,1950 年代后大约每 10 年增长一倍。第二,文献新陈代谢加快,老化周期缩短。据统计,目前各类文献的平均寿命是:图书 10~20 年,科技报告 10 年,学位论文

5~7年,期刊论文 3~5 年,技术标准 5 年,产品样本 3~5 年。面对浩如烟海的新知识、新情报,科技工作者只有望洋兴叹,盼望有一种手段能迅速、准确地从知识大海中获取他们所需要的那一滴。人们的检索要求也在不断提高,开始是通过浏览查阅资料;之后是希望文献工作者编制目录索引使他们查得更快一些;计算机出现以后,人们又指望用机器代替手工查找;随着计算机智能功能的不断加强,人们又要求除了简单匹配查找外,计算机把内容的判断识别也担负起来。

计算机软硬件技术的不断发展使以上希望成为现实。随着计算机匹配和推理能力的不断加强,内外存容量的不断扩大,数据库技术和通讯技术的发展以及操作系统的不断完善,计算机信息检索系统在人类社会生活和经济生活中的地位也日趋重要。

1.2.1 国外发展情况

计算机信息检索始于 1960 年代,它的发展可分为四个时期:

第一阶段(1964~1969 年):脱机批处理信息检索,即根据用户要求与说明,编写成逻辑提问式,在系统中建立户头,系统定期对新到资料按用户要求进行检索,将检索结果分发给用户。这一阶段的特点是“单机处理、顺序查找”,即用户直接利用一台计算机对单个数据库进行查找,而且由于数据载体是磁带,只能进行顺序查找。

第二阶段(1964~1970 年):联机信息检索,即用户直接通过线路一边输入信息,提出问题,一边处理信息,最后的检索结果也通过通讯线路返回到用户。我国现在的计算机信息检索系统大都

处于第一、第二阶段。这一阶段的特点是“多机局域网,倒排查找”。由于数据载体改为磁盘,用户可以随时提问,随时获得结果。

第三阶段(1970~1990年):大型联机信息检索,即多个用户终端共享一条线路与计算机相连,计算机通过远程分时处理系统和远程实时系统以联机方式开展脱机批处理服务,达到全球资源共享,著名的美国 DIALOG、ORBIT、DMS/ONLINE 和欧洲的 ESA-IRS 系统都处于这一阶段。这一阶段的特点是“远程通讯、大型机、多用户处理的操作系统”,即中央处理机为大型机,它具有多用户处理的操作系统,通过卫星通讯与世界各地检索终端相连。

第四阶段(1990~2000年以后):人工智能信息系统,即智能机器人和专家系统,它不仅能使用户从计算机中检索出信息,而且还可以帮助用户进行逻辑推理,启发用户思路,用户可以用各种自然语言进行检索。

目前,世界上提供全球性联机检索服务的机构通常是商业性私人机构,也有少量的政府机构和大学。这些机构提供检索用的大型计算机,培训用户,代查原始文献,还作为代理商安排用户和计算机之间的对话。至于供检索用的数据库,多半是由另外的组织(私人公司、政府机构、专业文摘社)生产的。检索服务机构以租用或交使用费的形式使用这些数据库。现在世界上最著名的国际联机检索服务机构有:美国的 DIALOG(洛克希德情报系统)、ORBIT(系统开发公司)、MEDLINE(国家医学图书馆)、欧洲航天局的 ESA-IRS 等系统,其中 DIALOG 系统目前已拥有 600 多个

可供检索的数据库,几乎涵盖了所有学科行业。

1.2.2 国内发展情况

我国计算机信息检索的发展始于 1975 年,大体经历了三个阶段。

第一阶段是萌芽阶段(1975~1980 年),特点是引进与试验。1975 年,国防科委情报所为实现自动检索和汉字处理,引进了日本高千穗 4100 汉字处理系统,这是我国信息自动检索的首次举动。当时我国还没有一个自制的汉字处理系统,中文文献的处理根本无从谈起。同年,中国科技情报所和北京图书馆发起成立《汉语主题词表》编辑组,共 500 多个单位的 1300 多人参加。三年后出版了综合性的《汉语主题词表》,为建立全国汉字标引和检索系统打下了基础。也是在 1975 年,北京文献服务处引进了“美国政府研究报告”(GRA)磁带。首次在国内开展了计算机检索定题服务。随后,机械工业部情报所、化工部情报所、南京大学、地质情报所相继引进英国《科学文摘》(INSPEC)磁带、美国《化学文摘》(CA)、《生物文摘》(BA)磁带等开展了类似服务。

第二阶段是生长阶段(1980~1990 年),以服务和自己研制为特点。1980 年,我国 11 个部委情报所联合在香港建立国际联机检索终端,首次在国内开展了国际联机检索服务。1983 年,中国科技情报所在引进的日本 TK—70 型计算机上实现了中国药学文摘自动编辑和汉字自动检索,这是我国第一个自己研制的中文文献自动检索系统。1983 年,中国科技情报所、水电部情报所、兵器工业部情报所在北京设立了国际联机检索终端,至 1986 年,这类