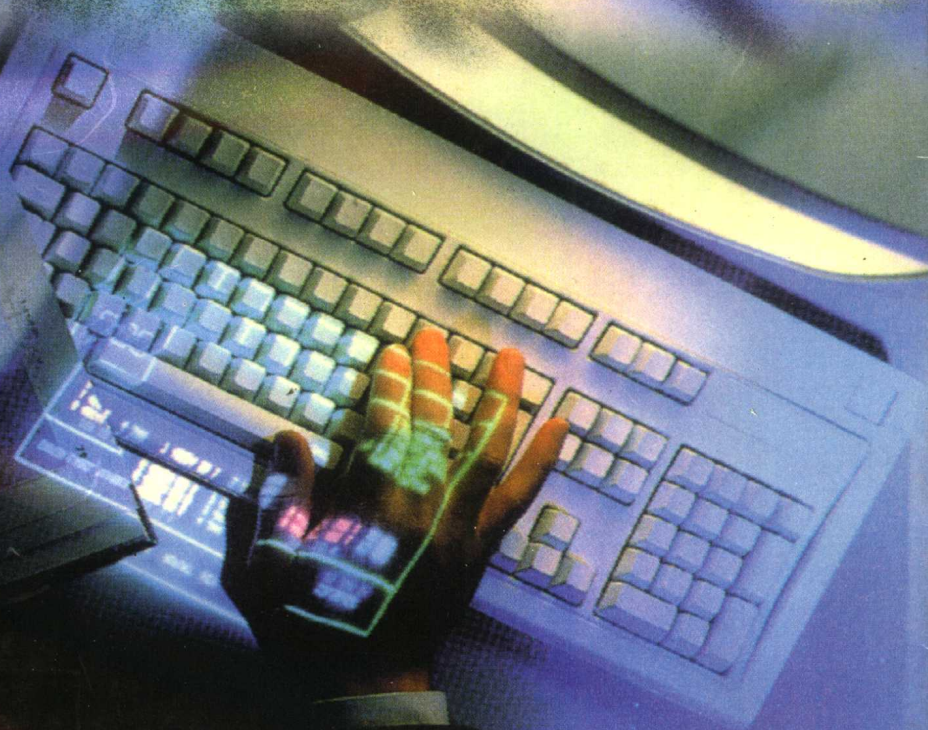


计算机情报检索

慎金花 何惠芬 陈欣 编著



36

同济大学出版社

计算机情报检索

慎金花 何惠芬 陈欣 编著

同济大学出版社

内容简介

本书比较详细地介绍了计算机情报检索的概况、原理和基本方法,计算机情报检索数据库的组织和结构以及一些著名国际国内联机检索系统的检索方法等知识。本书主要内容有:计算机情报检索概况;计算机情报检索数据库的组织和结构;计算机情报检索的基本过程与检索技巧;DIALOG,STN,BDS等国内外联机检索系统的检索方法;光盘检索系统的检索方法以及 Internet 信息检索等。

本书的读者对象为图书情报单位的工作人员,图书情报专业的学生,高等院校的本科生、研究生以及对计算机信息检索有兴趣的人员等。

责任编辑 张智中

封面设计 潘向蓁

计算机情报检索

慎金花 何惠芬 陈欣 编著

同济大学出版社出版

(上海四平路 1239 号 邮编:200092)

新华书店上海发行所发行

同济大学印刷厂印刷

开本:850×1168 1/32 印张:9.5 字数:270千字

1998年9月第1版

1998年9月第1次印刷

印数:1—1500

定价:15.00元

ISBN7-5608-1971-0/TP·213

前 言

计算机情报检索是本世纪 50 年代出现的新技术。40 多年来,随着计算机技术和通讯技术的不断进步,计算机情报检索在理论和实践两方面都得到了迅速的发展,为科学技术和社会经济的发展作出了重要的贡献。目前,世界上国际联机情报检索已发展到相当高的水平,联机网络和检索终端几乎遍及世界所有的国家和地区。我国在这方面虽起步较晚,但近十多年来发展很快。国内自建的科技情报联机检索系统也已初具规模,并投入了商业性运营。近几年来光盘网络和 Internet 技术的发展,使得计算机情报检索又进入了一个崭新的发展阶段,从事情报检索和信息服务的人数不断增多。

同济大学图书馆于 80 年代末开始面向研究生开设“计算机情报检索”课程。经过多年的教学实践,并参考了国内外一些有代表性的教材和大量文献资料,我们编写了本书。本书从用户使用的角度来介绍计算机情报检索的原理和方法,并且结合具体检索实例,力求做到重点突出、浅显易懂、实用性强,尽量不涉及过于深奥的计算机情报检索专业知识,使用户能够很快掌握计算机情报检索的基本方法和技巧,并熟练地应用于学习和科研之中;本书在完整系统地介绍计算机情报检索的基本原理和基本知识的基础上,尽量使用最新的数据和资料,并且加入了这些年来飞速发展的光盘检索和 Internet 利用等内容,使用户能够跟踪并掌握当今世界最新的计算机情报检索技术。

本书共八章,其中,第一、二、四章由慎金花编写,第三、六章由何惠芬编写,第五、七、八章由陈欣编写。在编写过程中,得到了肖友瑟教授和周维德副教授的热心帮助。

由于作者水平有限,书中难免有疏漏之处。敬请各位读者批评指正。

作 者
1998 年秋于同济

序

面对知识经济的大浪潮,国家正在落实科教兴国战略及知识创新工程,这对今天的大学生、研究生——新一代的重要科技力量来说,无疑既是大好的机遇,又是严峻的挑战。

早在三四十年前,一些卓有见识的人士,面对高科技、电子信息技术的发展,提出了知识将成为新的战略资源的论述,今天,这一论述已广泛地为人们所接受。也就是说,原来以物质形态作为主要资源来推动经济发展的时代将成为过去,代之以知识(智力)形态作为主要资源而推动经济的发展。这就要求我们在贯彻科教兴国、人才培养的工作中,必须紧跟时代步伐,立足知识创新工程。要创新,除必须具有扎实的专业基础知识及相应的创新精神外,还必须发掘新的起点及掌握如何获得新的起点的技能,从而在新的起点上,开拓思路,有所发明,有所创造。通过先进的信息技术,对学科领域进行检索查新,是立足学科前沿的必由之路。为此,国家教育部早就发出了“在高等教育中加强文献信息教学”的指示,同济大学科技情报研究所的老师们为贯彻落实这一指示,在开设《文献检索与利用》课程的基础上,针对当前电子信息技术的发展,对研究生开设了《计算机情报检索》课程,取得了良好的效果。编者积多年来从事这一课程教学的实践经验,汇集国内外在这方面的最新成果,在自编讲义的基础上,精心补充修订而成此教材。全书结构严谨,层次清楚,文辞通畅,深入浅出,在试用过程中,受到了读者好评。

本教材具有以下特点:① 在介绍一般基本原理与方法的基础上,突出实用性。考虑到学生学习这一课程的目的,在于掌握实际操作使用的能力,而不必像专业情报人员那样需要编程建库的技

能,因此取材充分注意了这一尺度,并注重联系工程实际举例,以便于学生掌握;② 在介绍一般计算机检索知识的基础上,突出新颖性,如注重引述了最新发展的 Internet 网络及光盘系统,以便学生掌握最新的一些检索途径,这些 90 年代以来发展的新技术,在目前已出版的一些同类教材中,因受编写年代的限制,尚较少涉及;③ 恰当地处理了重点突出,照顾全面,即点深面广的关系。如在联机检索中,在以使用最广泛的 DIALOG 系统为重点的基础上,也注意了 STN 等系统的介绍,同时对国内联机,也作了相应地恰当地引述。在触类旁通,以点代面方面做了较好的处理。

在本书中实例的具体指引下,配合适当的实习环节,更便于学生掌握所需的基本机检知识和技能。

肖友瑟

1998 年 6 月

目 录

第一章 计算机情报检索概述	(1)
第一节 计算机情报检索的产生背景及其发展过程	(1)
第二节 国际国内联机检索发展现状	(5)
第三节 联机情报检索系统的软硬件构成	(7)
第四节 联机情报检索的服务方式和优缺点	(10)
第二章 计算机情报检索系统数据库	(14)
第一节 数据库基本知识	(14)
第二节 计算机情报检索系统数据库的类型	(16)
第三节 计算机情报检索数据库的组织与结构及检索原理	(27)
第三章 计算机情报检索的基本过程与检索技巧	(37)
第一节 计算机情报检索的基本过程	(37)
第二节 计算机情报检索策略编写实例	(62)
第四章 DIALOG 国际联机检索系统	(65)
第一节 DIALOG 系统概况	(65)
第二节 DIALOG 系统常用数据库简介	(67)
第三节 DIALOG 系统检索方法与命令	(74)
第五章 STN 国际联机检索系统	(89)
第一节 STN 系统概况	(89)
第二节 STN 数据库	(90)
第三节 STN 系统基本检索方法与指令	(94)
第六章 国内联机检索系统	(107)
第一节 中国科技信息所联机系统 (TRIP 系统)	(107)
第二节 北京文献服务处联机检索系统 (BDS 系统)	(112)

第三节	上海科技情报所、机械部信息所、原化工部信息中心联机系统	(128)
第七章	光盘检索系统	(134)
第一节	概况	(134)
第二节	KR INFORMATION OnDisc 光盘检索系统	(136)
第三节	ProQuest 光盘检索系统	(143)
第四节	其他光盘检索系统	(146)
第八章	国际互联网络 Internet	(151)
第一节	Internet 概况	(151)
第二节	Internet 提供的主要服务	(157)
附录一	DIALOG 系统数据库分类表	(162)
附录二	STN 系统数据库分类表	(281)
附录三	DIALOG 系统文献型数据库蓝页范例	(286)
参考文献		(294)

第一章 计算机情报检索概述

电子计算机技术应用到文献检索上来,只有 40 多年的历史。所谓计算机情报检索,简单来说就是利用电子计算机对信息和数据的高速处理能力来存贮数据并从中查找、取出数据这样一个过程。这种现代化的情报检索技术的出现是人类知识急剧增长和现代化科学技术发展的必然结果。

第一节 计算机情报检索的产生背景 及其发展过程

一、产生背景

1. 文献数量急剧增长

当今世界的发达国家正处在从工业化社会向“后工业化社会”和“信息化社会”过渡的时代。信息已成为社会的一种重要资源,成为社会生产活动、科技活动、乃至日常生活中不可或缺的要害。近三十年来出现的科技成果,其数据超过了在此之前人类历史上两千年来的总和。据报道,当今世界每一分钟产生一本图书,每 30 秒产生一篇专利文献,每 5~6 秒产生一篇论文。目前,全世界每年出版的各种图书可达五六十万种之多,每年发表的国际科技论文可达十多万篇,申请专利一百多万件,科技论文五百多万篇,所有这些情况,使得文献种类和数量爆炸性地增长,由此,也给科研工作者阅读资料带来了不便。

2. 文献分布异常分散

分支学科和边缘学科的不断繁衍,使各学科间交叉渗透现象

日益突出。据统计:一个传统学科应用的技术 50% 是从别的学科来的,一个新兴学科应用的技术有 70% ~ 80% 是从外学科来的,由此造成学科之间在内容上相互交叉渗透,同一专业的文献往往分散在许多不同的刊物上。美国麻省理工学院曾经统计,就电工方面的文献而论,仅有 50% 刊登在 1000 多种电工杂志上,而其余的 50% 则分散在物理、机械、化工、生物等杂志上。这种状况给搜集情报资料、查阅文献的工作带来了极大的困难。在这种情况下,如果还用传统的手工方式在浩如烟海的文献资料中查找所需的情报,就会像大海捞针一样困难。

3. 文献寿命明显缩短

伴随着科学技术日新月异的发展,每时每刻都有新的发现、发明、创造和革新。反映在科技文献上就是文献的使用时间在缩短,旧的文献很快就失去时效而被新的文献所代替。据前苏联《发明问题》杂志的统计分析认为,各类文献的平均时效是:图书 10 ~ 20 年,科技报告 10 年,学位论文 5 ~ 7 年,期刊与连续出版物 3 ~ 5 年。而西方国家认为,80% ~ 90% 的文献的有效周期只有 5 年。

面对以上现实,如何提高人们准确、迅速地识别、搜集、处理、吸收情报和有效地利用情报的能力,已成为一个十分突出的问题。计算机技术、高密度海量存贮技术以及现代通讯技术的飞速发展,为人们利用计算机等现代技术来进行情报存贮和检索提供了物质前提。

电子计算机的出现与发展,为情报的自动存贮和管理提供了坚强的物质基础,网络、数据传输等技术又为远距离检索提供了有利条件,使联机检索成为可能,而国际卫星通讯技术的出现,使完全超越地区限制的“国际联机情报检索”变成了现实。所有这些外部条件与情报检索本身的发展要求相结合,就导致了计算机情报检索的产生和迅速发展。

二、发展过程

从 50 年代初计算机首次应用于情报检索试验,到如今的多元化全面发展,计算机情报检索经历了以下四个阶段:

1. 脱机检索阶段(1954 ~ 1964)

50 年代初、中期,由于计算机尚处于第一代电子管时期,用于情报检索有很大的局限性,因此几乎没有实用系统,有的也只是内部的实验性或半实验性系统,而且只能脱机检索。据记载,最早的计算机情报检索系统雏形是美国海军军械中心(NOTS)于 1954 年研制的。到了 50 年代末和 60 年代初,由于第二代计算机的大量涌现,计算机系统的软硬件都有很大发展,使得计算机文献处理与情报检索有了很大的活力,从而进入实用化的脱机批处理检索方式阶段。1961 年,美国化学文摘社第一个公开将计算机应用于文献型情报的处理中。1963 年美国国家医学图书馆也开始使用计算机来编制世界医学文献的检索刊物《医学索引》的月度主题索引。

由上可见,脱机检索阶段主要是利用计算机来进行文献目录、索引、文摘的前处理工作,也进行一些脱机定题检索。

2. 联机检索阶段(1965 ~ 1972)

这一阶段,由于第三代集成电路计算机的产生与发展和高密度海量存储器——硬磁盘及磁盘机的出现,再加上数字通讯技术的发展和分组交换公用数据通信网的普及,使得情报检索从脱机批处理阶段进入联机检索阶段。单台计算机通过通讯线路可联接多个检索终端,利用分时技术,若干个用户可同时直接和主机以“对话”方式进行检索。在这一时期,比较有代表性的检索系统有:美国洛克希德导弹与宇航公司下属的一个实验室于 1965 ~ 1966 年间建立的 DIALOG 系统,紧随其后的是美国国家医学图书馆的 MEDLINE 系统。由于这一阶段的计算机网络主要是通过电话线路联结,因而,联机检索受到地区的限制。

3. 国际联机检索的发展与普及阶段(1973 年至今)

这一时期,由于第四代和第四代半计算机的出现,以及卫星通讯技术和光纤通讯技术的实用化,使得计算机情报检索冲破时间和空间的限制而进入国际联机检索阶段。这种检索技术的出现,使更多地区的用户能够直接进行联机检索,大大方便与加速了世界情报资源的交流与利用。过去千里迢迢不能找到看到的情报资料,现在利用国际联机情报检索终端,只用几分钟就可以查遍几年到几十年的世界各国几万种期刊中的有关资料。国际联机情报检索技术的成功应用,为快速全面地获取全球性科技情报资料和经济信息提供了非常方便的条件,从而极大地提高了情报资料的可获得性和利用价值,充分实现了人类情报资源的共享。

4. 光盘检索系统的出现(1985 年至今)

联机检索发展到 80 年代,随着社会对情报需求的日益增长,数据库生产和服务的市场越来越大。1985 年,出现了 CD-ROM 数据库,CD-ROM 只读光盘数据库作为一种新的和重要的数据库存取介质,其发展出现了惊人的增长。它只需要一台微机,一个驱动器,一张 CD-ROM 盘即可运行。由于其使用方便,存贮量大,不受检索时间、通信费用、打印篇数的严格限制而深受读者欢迎。尤其是在目前国际联机检索费用较高的条件下,光盘检索对用户的吸引力更大。它只要一次性地投入成本(当然成本也较高),就可以反复使用。用户使用光盘来检索既可以反复修改检索策略以保证检索效果,又可以节约费用。光盘检索的出现使得计算机情报检索的内容更加丰富,服务方式也更加多样化,它与联机检索互相补充又各有优缺点。在实际检索过程中,只有将两者结合起来,相互取长补短,才能得到满意的检索效果。

由于光盘检索的硬件要求和服务方式等与联机检索有所不同,本章以下几节仅介绍传统的联机检索的有关知识,光盘检索作为一种新的计算机检索方式,我们将在第七章详细介绍。

第二节 国际国内联机检索发展现状

一、国际现状

目前世界上国际联机情报检索已发展到相当高的水平,联机网络和检索终端几乎遍及世界所有的地区和国家。总的来说,其发展有以下特点:

1. 开展服务的系统数量增加极快

目前开展国际联机数据库检索服务的机构已达 200 家左右,其中属于商业性的规模较大的著名国际联机系统有以下几家:

(1) DIALOG 系统。由 DIALOG Corporation 经营,1998 年拥有 900 多个数据库,文献专业范围涉及所有的学科领域,用户遍及 120 多个国家和地区的 2 万多个机构。

(2) BRS 系统。由美国文献题录检索服务公司经营,现有 100 多个公用数据库和 40 多个私人数据库,报道范围涉及各个学科,尤以产品信息和工业标准与技术规范方面的数据库独具特色。另外,它的全文检索服务优于其他联机系统。

(3) ORBIT 系统。原属美国系统发展公司,现是法国 Telecom Multimedia 的子公司,有 250 个公用数据库,以专利、商标、化学、科技等方面的数据库最具特色,其用户终端约 35000。

(4) ESA/IRS 系统。由位于意大利弗拉斯卡蒂的欧洲航天局情报检索中心经营,有 120 多个数据库,涉及各个科技领域,用户终端遍及世界许多国家。

(5) STN 系统。由美国化学文摘社,德国卡尔斯鲁厄能源、物理、数学情报中心以及日本科技信息中心三家联合经营,约有 150 多个科技方面的数据库,该系统以化学、建筑、纺织与能源专业数据库最具特色。

2. 联机数据库规模和数量迅速增长

目前,不仅国际联机检索系统日益增多,其功能越来越完善,

而且联机数据库的规模和数量也迅速增长。现在世界上绝大多数文摘和索引刊物都有机读版本,其中 80% 以上可在世界范围内供联机检索。而且由于纸张与印刷费用的不断涨价,没有对应的印刷版的电子数据库也逐渐增多。

3. 用户数越来越多

在许多发达国家,联机检索已成为获取科研资料的主要手段,其使用已相当普及。美国不少大学教授、公司经理和高级技术专家家里都设有联机检索终端。有人甚至将国际联机检索终端带到了海上钻井平台。

4. 服务领域越来越广

国际联机检索服务除了面向科技领域外,还面向经济、管理和社会科学甚至深入到社会公用事业服务与家庭生活的各个领域之中。联机检索服务已发展成为一种新兴的产业。它作为当今信息产业的一个重要组成部分,不仅产生了巨大的社会效益,同时也产生了巨大的经济效益。

二、国内联机检索概况

1. 国内利用国外联机检索系统情况

我国的联机情报检索起步较晚,国际联机情报检索又更迟一步。1980 年初,由当时的建工总局建筑研究院情报所等单位在香港设置了我国第一台国际联机情报检索终端,通过香港大东电报局与美国的 DIALOG 和 ORBIT 系统联机,从此开始了国内检索国际联机系统数据库的历史。

1981 年,石油部情报所和当时的北方科技情报所分别于 3 月和 12 月先后在北京与美国的 DIALOG 系统,ORBIT 系统直接联机。

到 1982 年底,全国已有北京、南京等五六家单位与国外著名联机检索系统联机。但由于他们采用的是 50 比特(bit)传输速率的电传终端,因而,耗费通信时间和机时较多,费用较贵,从而影响了用户的使用率。

1983年,中国科技情报研究所通过ESANET网开始与欧洲宇航局情报检索服务部(ESA-IRS)及DIALOG系统、ORBIT系统联机。

随后,中国科技情报所、北方科技情报所、水电部情报所等通过终端分机时的办法为本系统单位设立了国际联机分终端。发展到如今,在全国范围内已有几百家单位成为国际联机用户,一般的省市科技情报所及重点高等院校都有国际联机检索服务。而且,由于通讯技术的不断发展,数据传输速率已提高至64K或更高的速度,国际联机检索作为国内科研人员获取国外科技信息的一种手段已不再像以前那么神秘了。人们已经慢慢地接受并喜欢上了这一现代化的检索手段。

2. 国内联机检索系统的发展

除了利用终端检索国际联机系统以外,国内一些基础较好、实力较为雄厚的单位也开展了联机检索系统的开发和建设工作。北京中国科技情报所、机械部情报所、国防科工委情报所等单位都相继建立了自己的数据库和联机检索系统,并开始对外服务。目前,国内联机系统较大的有中国科技信息研究所、原机电部科技信息研究所、国防科工委北京文献服务处、原化工部科技信息所、原冶金部科技信息所、上海科技情报所等。

第三节 联机情报检索系统的软硬件构成

联机情报检索是借助检索终端通过通讯网络与联机情报检索中心的中央计算机联机来进行情报检索的。因此,一个联机情报检索系统应该包括检索终端、通讯系统以及联机情报检索主系统三个部分。

一、检索终端

它是用户与情报检索主系统的中央计算机进行“人机对话”的

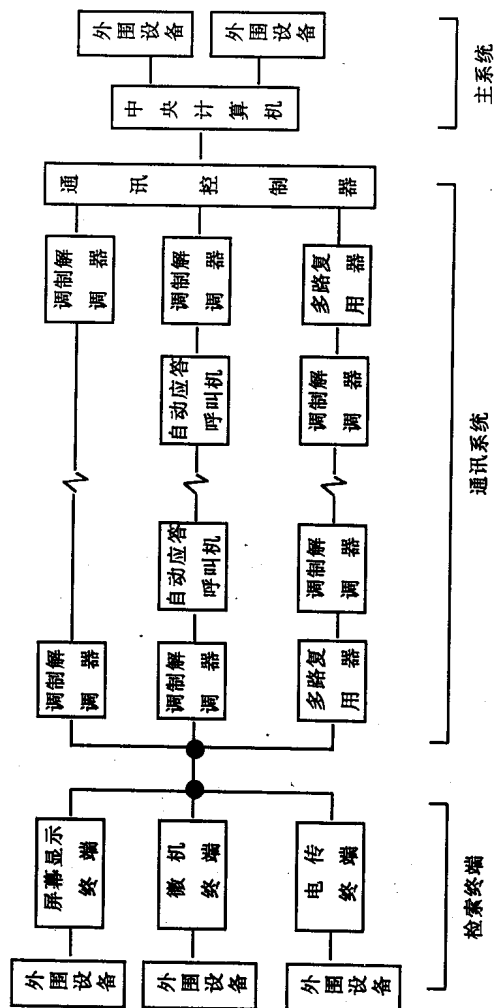


图 1-1 联机情报检索系统构成示意图