

光盘数据库 检索指南

毛文敏 黄筱刚 何 清 编著



上海科学技术文献出版社

光盘数据库检索指南

上海科学技术文献出版社

光盘数据库检索指南

毛文敏 黄筱刚 何 清 编著

上海科学技术文献出版社

(沪)新登字 301 号

光盘数据库检索指南

毛文敏 黄筱刚 何 清 编著

*

上海科学技术文献出版社出版发行

(上海市武康路 2 号 邮政编码 200031)

全国新华书店经销

上海科技文献出版社昆山联营厂印刷

*

开本 787×1092 1/16 印张 14.25 字数 356,000

1996 年 1 月第 1 版 1996 年 1 月第 1 次印刷

印数: 1-4,000

ISBN 7-5439-0536-1/Z · 659

定价: 22.00 元

前言

光盘技术是激光、计算机、数字通讯和光电集成等现代高新技术的结晶,在图书馆情报学界,光盘技术的渗透,集中体现在光盘数据库的涌现以及以此为核心的光盘检索系统的不断发展。光盘技术的应用为解决日益严重的情报增长与情报需求之间的矛盾,提供了有效的方法和途径。

《光盘数据库检索指南》一书较为详尽地介绍了目前国内利用较广泛的10种光盘数据库。它以论述各种光盘数据库的结构和检索方法为核心,列举了大量的应用实例。在论述中,力求理论联系实际,深入浅出,侧重于实用。

本书共分十章:第一章、光盘数据库检索基础;第二章、SPIRS检索系统;第三章、Silver Platter光盘数据库;第四章、《美国专利》光盘数据库;第五章、《美国工程索引》光盘数据库;第六章、《化学文摘》光盘数据库;第七章、《高技术文摘》光盘数据库;第八章、《美国公司样本集》光盘数据库;第九章、《美国军用标准》光盘数据库;第十章、《中文科技期刊篇名》光盘数据库。其中,第一、二、三、四章由毛文敏执笔编著,第五、六章由何清执笔编著,第七、八、九、十章由黄筱刚执笔编著。

本书在编写过程中,得到了邹乔敏同志的热情支持和帮助。顾耀芳承担了部分翻译工作。本书的部分审稿工作由吴建新担任。另外,还得到了王孟官、胡飞珍、梁建新、朱普德四位同志的多方帮助,在此一并致谢。

光盘技术的发展和光盘数据库的普及是迅速的,本书作者谨以此“指南”作为有关这一领域的第一本具有实际意义的操作性手册问世社会,期望接受同行、专家们的鉴评和实践的检验,以求更臻完善。

编著

ADR-1005

内 容 提 要

本书介绍了光盘技术在图书馆、情报界的应用,并对“美国专利”、“化学文摘”、“美国工程索引”等多种光盘逐一进行系统地分析阐述,较全面地反映了目前国内利用较广泛的光盘数据库的情况。

本书以论述光盘数据库的结构和检索方法为核心,以介绍各种检索字段为重点,突出理论联系实际,列举应用实例,侧重于实用。

可供情报检索管理人员、科研人员、大专院校师生使用。

目 录

第一章 光盘数据库检索基础	(1)
第一节 光盘和光盘数据库	(1)
第二节 微机光盘检索系统	(5)
第三节 光盘检索的特点和作用	(8)
第四节 CD-ROM光盘数据库检索提问表达式	(9)
第二章 SPIRS检索系统	(14)
第一节 SPIRS的安装和启动	(14)
第二节 SPIRS的结构及其主要检索功能	(22)
第三节 SPIRS的命令功能	(30)
第三章 Silver Platter光盘数据库	(39)
第一节 LISA光盘数据库	(39)
第二节 MEDLINE光盘数据库	(47)
第三节 GRADLINE光盘数据库	(51)
第四章 《美国专利》光盘数据库	(62)
第一节 《美国专利》光盘数据库介绍	(62)
第二节 《美国专利》光盘数据库各项检索功能	(66)
第三节 APS的各种功能键	(77)
第五章 《美国工程索引》光盘数据库	(82)
第一节 数据库内容简介	(82)
第二节 数据库的安装、检索	(82)
第六章 《化学文摘》光盘数据库	(106)
第一节 数据库内容简介	(106)
第二节 数据库的安装及种类	(106)
第三节 浏览型光盘	(109)
第四节 检索型光盘	(120)

第七章 《高技术文摘》光盘数据库	(132)
第一节 概况	(132)
第二节 安装	(132)
第三节 启动和退出光盘数据库	(133)
第四节 浏览索引(Browse)	(134)
第五节 检索(Search)	(135)
第六节 格式(Format)	(142)
第七节 功能(Action)	(144)
第八节 选择(Options)	(145)
第八章 《美国公司样本集》光盘数据库	(147)
第一节 概况	(147)
第二节 安装	(147)
第三节 启动和退出《美国公司样本集》光盘数据库	(148)
第四节 检索方法	(148)
第五节 打印和拷贝	(157)
第九章 《美国军用标准》光盘数据库	(159)
第一节 概况	(159)
第二节 安装	(159)
第三节 启动和退出《美国军用标准》光盘数据库	(160)
第四节 检索方法	(161)
第十章 《中文科技期刊篇名》光盘数据库	(172)
第一节 概况	(172)
第二节 安装	(172)
第三节 启动与退出光盘数据库	(173)
第四节 检索方法	(175)
第五节 显示、打印和拷贝	(179)
第六节 检索实例	(181)
附录	(183)
一 《国际专利分类简表》	(183)
二 《化学文摘》类目表	(201)
三 “联邦供应分类表”大类类目表	(202)
四 “美国情报处理服务公司”分类大类表	(204)
五 上海市主要光盘数据库分布概况	(217)

第一章 光盘数据库检索基础

光盘技术是激光、计算机、数字通讯和光电集成电路等现代高新技术的结晶。光盘存储器以其高性能、多功能、多类型等突出优点而成为继磁存储器之后更为新颖有效的现代化信息存储和传播工具,可以说,它的诞生和发展,在信息领域里掀起了一场新的革命。在图书情报界,光盘技术的渗透,集中体现在光盘数据库的涌现以及以此为核心的光盘检索系统的不断发展。光盘技术的应用为解决图书情报界所面临的日益严重的情报增长与情报需求之间的矛盾,提供了一种有效的方法和途径。本章首先对光盘技术、光盘数据库以及光盘检索系统的组成和使用作一概述。

第一节 光盘和光盘数据库

一、光盘技术概述

最早的光盘是1972年由荷兰飞利浦公司研制成功的激光电视唱盘,它用激光存储和读取电视图像信号,与普通录像磁带相比具有存储密度高、使用方便等优点,因而迅速打开了市场。不久,1978年飞利浦公司又推出了数字式音频唱片,也称激光唱片,它是用激光将数字化的音乐节目录入唱片,以其优美的音色而广受青睐。光盘产品在娱乐界的迅速崛起,使人们意识到光盘技术也可能在信息科学领域里引起一场新的技术革命。于是世界各国的著名公司都纷纷投入光盘存储器的研制,并将研究重点从民用音像领域转到计算机领域,以解决计算机用的磁介质存在着的不能长期保存、存储量有限的固有弱点。

80年代是光盘技术的研究和产品取得突破性进展的年代。自1983年首张高密度只读式光盘存储器(CD-ROM)诞生以来,光盘作为计算机的外存储器引起国际信息工业界的极大兴趣。据统计,从1985年到1990年,美国CD-ROM光盘系统的总销售额由28.2万美元增长到12.83亿美元,增长了4000多倍;1989年CD-ROM驱动器发货量超过16万台,比1988年7.5万台增加了1倍多;1990年世界销售CD-ROM量达1230万张,其驱动器达53万台。这一切表明,光盘存储器作为新一代的电子存储介质已得到普遍接受和广泛应用。

尽管有各种各样的光盘,但究其记录和读取信息的原理是基本相同的。在记录信息时,利用激光器发出的激光束通过光学系统聚焦成直径在1微米以下的微小光斑,并使小光斑在光盘上作径向恒速运动,同时光盘安装在精密转轴支撑的转台上,使转速保持恒定。这样,在光盘旋转和小光斑径向移动的共同作用下,便在低熔点的光盘膜上逐点灼烧成由小坑和平面(代表一定的编码D0E和D1E)组合起来的螺旋线。在读取信息时,将激光束聚焦成同样大小的光斑,也是在光盘旋转和小光斑径向移动的作用下,根据光斑在凹凸不平的螺旋线上所

产生的反射光束的强弱变化,还原成原来的存储信息编码。正是利用这种光学记录和读取信息的原理,使得光盘存储器具有存储密度高、存储容量大、介质成本低、电子存储功能强、存档特性优良、数据可靠性高、介质寿命长等特点。

目前光盘存储器的种类繁多,且形成了产品系列。根据读写光信号的不同方式,目前世界上使用的光盘存储器大致可分为“只读型光盘”、“单写多读型光盘”和“复写型光盘”3种。

(一)只读型光盘(Read-Only Optical Disc)

只读型光盘是采用批量模压制造工艺生产的,它上面的信息是由生产厂家用激光一次性录入的。用户可反复使用盘内的这些信息,但无法更改这些信息。因此,只读型光盘产品必须要有相当的数据量和发行量,这样在技术上和经济上才较为合算。目前只读型光盘的主要产品有:

1.激光唱盘(CDA-Compact Disc Audio) 激光唱盘,也叫激光唱片,它用数字化编码表示音符,其驱动器可接立体声音响设备。激光唱盘,音质优美动听,立体感强,唱盘可以用微机进行控制,随机播放唱盘上的任何一段乐曲。这种唱盘是目前世界上使用最多的一种光盘。

2.激光视盘(CDV-Compact Disc Video) 激光视盘也称影视光盘,是影视录像盘的变种,它采用模拟方式记录音频、彩色图像等信号,也能存储数字数据供微机使用;与CD-ROM光盘相比,它能把声、像、文字和颜色集于一体,是一种较高级的存储介质。激光视盘应用范围很广,也可用于彩色文献资料的存储与检索等领域。

3.高密度只读存储器(CD-ROM-Compact Disc-Read Only Memory),从本质上讲,CD-ROM是激光视盘的变种,与激光视盘相比,它是存储数字数据比较有效的存储介质,因为它设有将模拟信号转换成数字信号的过程。CD-ROM光盘主要用于文字、声音、图表、静止画面或软件等信息的存储,它以极高的存储密度、低廉的成本价格而成为计算机的一种海量外存。并且,它也是电子出版物的主要介质,广泛应用于图书情报领域和商业界。本书所述的光盘数据库实际上就是CD-ROM产品。

4.内联式只读光盘(CD-I-Compact Disc-Interactive) 内联式只读光盘,也称交互式光盘,或多功能光盘,它既可存储文字,也可存储图像、声音等多种信息形式,并且存储内容可通过软件进行对话。

5.扩展只读光盘(CD-ROM/EA-Compact Disc-Read Only Memory/Extended Architecture) CD-ROM/EA可以使CD-ROM扩充有CD-I的功能。CD-ROM/EA的信息除文本外,还附有声音、图表和影像,它既可以在CD-ROM工作站上放演,又能在CD-I放演器上使用。因此,CD-ROM/EA是一种在CD-ROM和CD-I之间起桥梁作用,并使出版商能生产出在两种系统上都能使用的多媒体。

6.内联式数字视盘(DVI-Digital Video Interactive) 内联式数字视盘,又叫交互式数字视盘,它在CD-ROM光盘的基础上,增加了存储活动图像、三维图形、多声道音乐等功能。

7.光学只读存储器(OROM-Optical Read-Only-Memory) 光学只读存储器的主要特点是采用同心圆(CAV)的存储方式,而不是光盘通常采用的螺旋式(CLV)存储方式,据称这样可获得较短的数据访问时间。

(二)单写多读型光盘(WORM - Write - Once Read - Many Optical Disc)

WORM本身是一个主盘(Master Disc),它已被事先格式化并隐含有驱动程序,用户可使用空白磁道写入自己的数据资料。启用时,WORM光盘系统的用户需要一套读写设备,该读写设备应该具有高能刻写和低能读书这两种激光头,或者是有1个可调激光能量的激光头。但在WORM上,数据一旦写入,便不能修改与擦除,只能多次读取。从计算机的外部存储介质的角度来看,WORM光盘很可能是一种过渡性产品,它有可能被复写型光盘所替代,但它在某些需要严格保证数据原样、绝对不能改写的应用领域,生命期才刚刚开始。可以预言,WORM可能在图书情报资料存储、后备存储等方面,以及军事、经济、财务、保密等领域里大有用武之地。

(三)复写型光盘(Multi - Write Optical Disc)

复写型光盘,也就是D可擦式光盘E(Erasable Optical Disc),是第三代光学存储介质。复写型光盘的反光材料已磁性化,因而它象当前广泛使用的磁盘那样可任意写入数据和改变原数据,据称,可提供用户重写万次以上。另外复写型光盘的存储量大(1张复写型光盘存储器可达650MB),且使用寿命长(一般在10年以上)。由于复写型光盘的这些优点,它是计算机尤其是微机的海量外存的有力竞争者,从总体上说,它取代磁盘作为微机的主要存储器将是一种发展趋势。

对于上述三代光盘存储介质来说,单写多读型和复写型光盘还未真正进入实用阶段,而只读型光盘则已发展得相当成熟并得到极其广泛的应用。特别是其中的CD-ROM光盘具有数据存储量大、使用寿命长、电子化处理功能强、价格低廉、维护方便等特点,在图书情报部门的档案、图像、文献资料等的海量存储、传播和利用等方面发挥着重要的作用,主要表现在CD-ROM光盘数据库的大量涌现以及光盘检索系统的不断发展。

二、CD-ROM光盘数据库

现代社会是信息的社会,信息量在不断激增,且包罗万象、瞬息万变,并与人们的社会生活和生产息息相关。可以说,谁获得信息最迅速、最丰富、最准确,谁就取得了一把打开成功大门的金钥匙。信息在未经处理时总是零星的、单一的、小范围的,很难满足人们的“快、多、准”的要求。因此多少年来,人们在不断地探求和发展有效的信息处理、存储和传播手段。70年代初的国际联机情报检索系统的诞生,开创了海量信息自动化处理的新时代。然而,80年代中期CD-ROM光盘数据库的出现则更代表着信息处理技术上的惊人发展。

CD-ROM光盘作为大型脱机式数据库的主要载体,其优势主要有以下几个方面:

1.有极高的存储密度和存储容量(参见表1-1) 由于激光束能用透镜聚焦成直径只有 $1\mu\text{m}$ 以下的光斑,因而信息存储密度可达 $10\text{bit}/\text{mm}^2$ 。这样,每张光盘上的存储容量便高达550MB(实际容量为650MB,修正软件等要占据一定空间),为软盘的1600倍,缩微平片的1200倍,相当于约25万页A4复印纸的信息量。

2.只读特性、成本低廉、复制简便、易于传递,是理想的电子出版物介质 CD-ROM的只读性和不可更改性对于数据库的出版、发行和控制来说是极其适宜的。一般来说,数据库一旦产生,就有其相对稳定性,其更改是由数据库生产者进行的,使用者没有必要也不可能更改数据库中的内容。这样,采用CD-ROM作为数据库的存储介质,就能有效地维护数据库的完整性,避免数据库传递中的数据丢失和损坏。另外,CD-ROM的只读性,也意味着它

的造价低廉且可大量复制,同时由于它体积小,也便于象其它出版物一样通过邮局发行,所以是理想的电子出版物媒体。

表1-1 CD-ROM与其它载体的存储容量比较

CD-ROM	A4纸	扫描图像 300×300 dpi	5.25英寸软盘 360KB/张	缩微平片	数据传输 1200波特 (baud)		磁盘 40MB	标准磁带 16000BPI
(张)	(页)	(页)	(张)	(张)	(日)	(h)	(张)	(盘)
1	27.5万	20万	1500	1200	46	1104	14	12

3. 能与微机结合起来使用,情报的处理和检索,均可由电子系统自动控制完成
CD-ROM设计的目标主要是针对微机,它的性能价格比与微机相当。由于目前微机正以空前的普及率渗透到各个领域,凡具有微机的用户,只需花少量费用购置1台CD-ROM驱动器,就可在微机环境下运行较大型的数据库。在微机上实现大型机的容量,这在CD-ROM以前是很难实现的。所以,CD-ROM与微机的结合能力,也是CD-ROM得以迅速推广的重要原因。同时,借助于微机的优势,CD-ROM数据库中数据的处理和检索,均可通过微机由电子系统自动控制完成。CD-ROM数据库中的软件大都实现了“用户友好”,不仅软件功能较强,而且充分考虑了用户的方便与需要,使CD-ROM数据库用户成为真正的最终用户。CD-ROM在使用上的方便、灵活和较强的针对性、适用性,对于建立一种分散的、小型的、内部的、以介质交换为条件的情报系统来说,有着重要的意义。

4. 具有优良的存档特性 由于CD-ROM属非接触式读写,读写头与盘片表面之间有0.3~1mm的距离,而且盘片表面和铝反射层(真正的信息层)之间还有1mm厚的透明层,也就是说,读写头与信息层之间的距离在2mm以上,不存在机械摩擦;而且,盘片不易受灰尘、指纹等的影响,对温度、湿度、电磁场等环境条件要求也很低,因此具有较高的稳定性,用它保存信息可达10~20年之久,而磁盘仅为1~3年。另外,CD-ROM在制造过程中增加了校验码,它能使读出装置校验被读出的数据是否完整,当不完整时,可插入被遗漏的信息,及时纠正出错信息,这对存档来说也是很重要的,它提供了用户所希望的高可靠性。

正是由于CD-ROM光盘的上述优点,以及现代信息社会发展的迫切需要,使得CD-ROM数据库在问世不到10年的时间里,发展速度相当惊人。1985年,在国际联机检索系统出现15年之后,第一张正式的CD-ROM数据库产品——BIBLIOFILE(《美国国会图书馆机读目录》)诞生了。当时,国际联机检索系统已进入成熟期,联机终端遍及世界各地,而CD-ROM数据库仍很快发展成可与之平起平坐的竞争对手,在短短的几年里掀起了发行CD-ROM数据库的热潮。据统计,至1988年6月,国外光盘数据库产品达200余种,1989年为496种,1990年有861种,而到了1991年已达到1552种。同时,光盘数据库的类型也不断丰富,除了最初的书目数据库外,又增加了文摘数据库、数值事实参考库、全文库等等,并且还出现了图像型、音频型、软件型和多媒体型等多种新颖的CD-ROM产品。

与国外相比,我国光盘数据库的起步并不算晚,并已取得了可喜的进步。早在1986年4月,国家海洋科技情报研究所就首先在我国引进了CD-ROM光盘数据库ASFA(《水科学

与渔业文摘》)和LSC(《生命科学文摘》)。到1987年底,国内共拥有11套8个品种的CD-ROM光盘数据库。1990年,全国拥有光盘系统的单位达到54个,共有93套30个品种的光盘数据库,到1991年,我国有126个单位引进了53种199套光盘数据库。不仅如此,我国已研制成了中文CD-ROM数据库,例如中国科学技术情报研究所重庆分所研制的《中文科技期刊篇名库》,以及中科院上海有机化学研究所的《中国化学文献数据库》等。这些CD-ROM的发行、使用,已取得了很好的社会效益和经济效益,我们相信,还会有更多更好的中文CD-ROM产品不断问世。

值得一提的是,上海科学技术情报研究所目前是全国范围内光盘数据库收藏量最大的单位,共收有16种西文CD-ROM数据库和2种中文CD-ROM数据库(其具体品种详见附录五),并全面对外开展服务。

第二节 微机光盘检索系统

CD-ROM光盘是一种信息载体,要对其中的信息进行检索和利用还需要计算机的配合。前面已谈到,CD-ROM的设计目标是针对微机的,因此我们在此主要介绍一下微机光盘检索系统,并分硬件配置和软件配置两部分来介绍。

一、硬件配置

微机光盘检索系统是以微机、光盘驱动器、光盘为主要组成部分而建立起来的,其中光盘是信息的载体,光盘驱动器是读取光盘中信息的设备,而微机则是整个系统的控制中心,光盘系统的各项功能就是围绕微机的能力而建立和完善的。下面分别阐述这些配置的规格和技术要求。

(一)光盘

由于CD-ROM已是发展较为成熟的技术,因此在物理规格和存储方式等方面都已形成了一定的国际标准。现在通用的CD-ROM光盘是一种直径为4.72in.(120mm)、厚度仅1.2mm的塑料(聚碳酸酯)圆盘,中间有一个直径为0.59in.(15mm)的圆孔。该塑料圆盘的一面是凹凸不平的、带有由许多凹陷的小坑组成的图形(代表一定的信息),在这一面的小坑图形上覆盖了一层金属铝膜(反射层),为了保护这层金属铝膜,在其表面涂上了透明的胶体;而读取信息则是在另一面,即透明的塑料一面。

另外,CD-ROM的存储方式主要有两种,即文本字符方式和图像传真方式。文本字符方式用于文字信息的处理相当有效,存储密度高,并可供检索以及转载、套录和编辑之用。图像传真方式则可处理各种文献资料(包括文字和图像),通过图像扫描仪存储信息,简便快捷,避免了繁琐的信息编码输入。这两种存储方式各自有特定的数据压缩技术,有时两种存储方式在同一盘片上结合起来使用,以达到更好的效果。图像方式的CD-ROM对微机的显示器要求更高一些,这在后面还将谈到。

(二)光盘驱动器及其控制板

光盘驱动器读出光盘中的内容并将它输入到计算机,控制板则是插在主机的扩展槽内,承担计算机与驱动器之间的连接任务,使计算机能识别驱动器提供的信息,并能实现对驱动器的控制。控制板可随光盘驱动器同时购买。

光盘驱动器的种类有:

1. 单机驱动器(Stand Along CD-ROM Drivers) 该驱动器与其自备的电源和机壳一起提供给用户,目前大多采用一个标准外部接口或者一个装在微机扩展槽内的专用设备接口卡与微机连接。这种驱动器的大小约与普通录像机相似,它的主要优点是:携带方便、连接简单、应用广泛。目前使用的驱动器主要是这种类型,其主要技术参数参见表1-2。

表1-2 单机驱动器主要技术参数(以日立CDR-1502S型为例)

转速:200-535 r/min (CLV)
容量:552MB
数据传递率:153KB/s
寻找时间(相邻轨道):1.2ms
平均取数时间:1s
软出错(可恢复): $<10^{-9}$
硬出错(不可恢复): $<10^{-12}$

2. 内藏式驱动器(Build in Drivers) 这种光盘驱动器象软盘、硬盘驱动器那样,装在微机内部。第一代内藏式驱动器按全高5.25in的标准设计,但最近有些型号的驱动器以半高形式装在更新更小的台式或便携式微机中,可随意搬动。

3. 多机驱动器(Multidrivers) 多机驱动器可将多个光盘驱动器连接、叠加在一起,其目的主要是减少换盘次数。例如,有4台驱动器相连,则一次可同时换4张光盘,但检索时仍须以1张光盘为单元,依次进行信息检索。所以,这种驱动器仍未根本解决扩大存储容量的问题。

4. 多盘箱式驱动器(又称投币式自动电唱机,Jukeboxes) 多盘箱式驱动器是目前市场上推出的一种最新型的驱动器。剑桥大学文摘社提供的多盘箱式驱动器内可装4张光盘,但只需要1个设备接口和相应的软件便可与计算机的操作系统和应用软件通信。它的优点是检索人员一次能检索多达4张盘的数据而无需频繁换盘或多次重复检索,因而检索速度快,但是这种驱动器的价格较贵。

近几年来,光盘驱动器的性能和可靠性不断提高。首先,平均存取速度已提高到300ms,达到相当于软盘的性能。在驱动器的结构上,采用自动透镜清洁装置、双门结构、密封机壳、长寿命马达,从而使无故障运行时间提高到2~3万小时。

(三)微机及输入输出设备

整个光盘系统的操作都是通过微机进行控制和执行的,并以微机的输入输出设备与外界进行信息交流。主机一般选择IBM-PC/XT或AT型微机(或IBM-PS2)或其兼容机。不同的CD-ROM产品对微机的内存可能有不同的要求,但至少要有256KB的内存,外存最好有20MB的硬盘。另外,带有图像的CD-ROM数据库对显示器有一定的要求,最起码应是VGA型的。其它如键盘、打印机等设备,CD-ROM数据库对它们没有特殊要求,在此就不赘述了。

二、软件配置

光盘系统的软件配置分两部分,即系统软件和应用软件。系统软件的主要功能是,使计算机可以运行和操作光盘驱动器以及识别光盘存储的信息。系统软件也是与光盘驱动器同时出售的。而应用软件则是与特定的光盘数据库相对应的,它使计算机能对特定的光盘数据库进行检索、套录、打印等操作。由于不同的光盘数据库一般要求有不同的应用软件,而应用软件的优劣又与光盘系统的功能密切相关,因此我们重点介绍一下应用软件的一些情况。

CD-ROM应用软件的基本功能有:(1)提供用户对CD-ROM文档进行读取的检索功能;(2)屏幕显示、键盘命令及查找与处理(打印、套录、编辑)等功能。可以看出CD-ROM应用软件的这两个基本功能保证了CD-ROM信息的有效检索,并且对用户是非常友好的。由于CD-ROM应用软件的主要功能是检索及其友好的用户界面,所以它又被称作检索软件。

CD-ROM检索软件与CD-ROM数据库通常是两个不同的产品,有时是由两个不同的公司生产的。检索软件程序既可放在CD-ROM数据库盘上,也可以独立的加密软盘方式提供给数据库使用者。目前,检索软件大多以后一种方式提供,这样既有利于软件修改,又不占数据库空间。另外在检索软件中,通常还有一部分安装程序,通过安装程序把检索软件安装到计算机硬盘的指定子目录上,以便于检索软件的运行使用。

目前,生产CD-ROM检索软件的厂家有很多,如Silverplatter(银盘公司)、DIALOG、UMI、Online等等,它们生产的软件,在检索、输出、显示和后处理功能上都是不一样的,它们拥有各自的产品系列。本书的后面几章将详细讨论一些主要的检索系统,在此,首先对各种检索软件所提供的检索模式和用户界面作一概述。

1. 引言性屏幕显示 这是用户进入光盘系统后第一眼就能看到的東西,它给出一些基本信息,如当前驱动器中光盘数据库的收录时间范围、数据库内容说明、系统情况及命令。有些系统还在屏幕下端给出用大写字母印出的当前功能,使用户知道系统正准备做什么。

2. 检索模式 光盘系统的检索模式主要有:布尔逻辑运算、截词检索、词典扩查、字段限定以及嵌套逻辑。但是在不同的光盘系统中,这些检索模式的应用形式是不同的,如有些系统要求用提问表达式输入,而有些系统则显示一个工作单,工作单上的各行设计成各个字段的专用输入行;又如有的布尔运算符是用简单的命令输入,有的则是用隐含的方式提供。不管如何处理,各系统一般都能做到输入要求清楚、命令简单、击键次数少。为此,有的系统还提供保留检索策略和重新执行检索策略的功能,并且修改和精炼检索策略的过程也是相当简单的。

3. 后处理功能 后处理功能包括检索结果的显示、打印、套录、重新确定格式及编辑、排序等功能,系统提供的选择项,可按用户的要求将检索结果在屏幕上显示、或打印成硬拷贝、或套录到软盘上。有的系统还允许用户提出输出的格式和排序的要求。

4. 屏幕帮助信息 光盘的用户界面是流动灵活的菜单式结构,有的菜单不仅可以作为其它菜单的子集出现,还可以从特定的命令级中产生。有时候屏幕上可以同时显示几个窗口,而新的窗口又可以响应用户的命令而出现。一般在当前菜单或窗口中显示调用其它相关窗口或菜单的命令,于是可以随心所欲地招呼各种功能。另外,在操作过程的任何时候,只要按“F1”功能键,就会立即看到有针对性的“帮助”屏幕,于是初学者也会感到很方便。

第三节 光盘检索的特点和作用

前面已对CD-ROM光盘数据库及其微机检索系统作了简单的介绍,在此基础上,我们将进一步分析光盘检索的特点和作用,并且主要通过光盘检索与传统手工检索及国际联机检索的比较来加以说明。

一、光盘检索与传统手工检索的比较

光盘检索与传统手工检索相比,不仅使用方便、检索功能强、输出灵活,最主要的,它还能节约大量的时间。由于光盘有极高的存储密度,1张CD-ROM光盘上平均可存储5年的信息量,而同样5年的工具书可能就要厚厚一大叠。据统计,CD-ROM光盘检索的整个过程平均约需20min,而人工查阅工具书则要花上1~2天的时间。印刷型工具书提供的检索途径和索引总是有限的,查找起来也比较繁,而CD-ROM光盘系统一般都具有组配检索、截词检索、位置检索、词典扩检、字段限制检索、分步检索等功能,因此可检字段多,组配相当灵活,检索效果好。另外,CD-ROM的检索结果可通过打印、拷贝等方式输出,并且能满足特定的排序要求和格式要求,而手工检索时最多只能得到零散的文摘复印件。

当然,光盘检索要收取适当的服务费,这也许增加了用户的经济负担。但是,按目前的收费标准,检索一个课题所需要的费用,一般在50~60元,相对于它的高效灵活的检索效果而言,这点经济投入可能就是必要的和合算的,而由此为用户节省下来的大量的宝贵时间更是无法用金钱来计算的。

二、光盘检索与国际联机检索的比较

相对于早已发展成熟的国际联机检索来说,CD-ROM也显示出强大的生命力和竞争力。联机检索的主要特点是数据库容量大,数据更新及时,检索响应时间短,这对于综合性的大课题检索、回溯性检索、实时性检索是比较适宜的。但是人们在适应了联机系统带来的种种好处之后,也开始抱怨起它的那些多年来一直无法改变的致命缺点,如对主机和通讯线路的依赖性,对用户的不友好,以及昂贵的费用。而CD-ROM光盘检索则在这些方面占有优势。CD-ROM光盘检索系统一旦建立,便成为一种不受外界条件(联机商、远程主机、通讯线路)影响的完整的检索系统,检索的主动权完全掌握在检索者手中,因而可以获得丰富的检索功能和理想的检索效果。

首先,CD-ROM光盘系统是自成一体的完整的检索系统。使用光盘检索系统时,用户不受通讯线路和主机的限制,这样就克服了通讯线路不畅、通讯费用高等不足,因而较国际联机检索更为方便。联机检索系统中全文数据库较少,因此即使检索到了所需文献,还须从纸件和缩微文档中获取原文,而CD-ROM系统则能允许我们在一种高性能载体上获得各种情报产品——用于检索的书目产品、用于全文检索的全文情报产品、可显示打印的图像情报产品等等。这样我们几乎能在检索的同时,获得和原文同样质量的复印件或软盘拷贝。这种将检索、获取原始文献、打印融为一体的完整检索系统使检索用户能对它行使最大限度的控制。

其次,CD-ROM系统使用户成为真正的检索者。联机检索时,要求采用较复杂的检索方法和命令,致使目前仍只有为数很少的真正用户能熟悉联机检索服务,它作为一种专门技术主要掌握在中介人手里。而在CD-ROM上进行检索是很方便的,用户很快就能熟悉其检

索方法,尤其是可以随时显示的帮助屏幕更使人感到清晰明了。CD-ROM这种“用户友好”的检索方式拓宽了CD-ROM的用户范围,使那些没有受过训练的情报用户能亲自参与或独立完成检索操作,培养了他们对CD-ROM情报源的亲切感,并能使他们意外地得到一些检索中介人发现不了的情报。

再次,由于在使用CD-ROM时,检索者可以从联机检索的钟点声和高额费用的压迫中解脱出来,因而能从容不迫地发挥多种检索功能和重复检索的优势,提高检索的效果。虽然CD-ROM检索软件所提供的检索功能与联机相似,但有些功能和命令在联机时因受时间限制而无法频繁使用,而使用CD-ROM时就可充分发挥其软件的优越性,对检索词反复核校,对检索策略不断加以完善,并能满足检索者在检索过程中临时产生的情报需求。另外,联机检索时出现的一个致命弱点是用户无法当场知道检索结果中到底有多少真正命中的文献,并常常要为大量的错误命中而付费。CD-ROM检索时就可逐条显示命中记录,从而能够清除错误命中,或进一步修改检索式以提高命中率。

总之,光盘检索的性能不比联机检索差,检索效果还优于联机检索,而其费用却只有联机检索的 $1/5 \sim 1/4$ 。虽然光盘检索的响应时间较长一些,数据更新较慢一些,但也都在一般用户可以接受的范围之内。尤其是在我国目前通讯设备比较落后、费用较高的情况下,光盘检索更显示出其优越性,从而成为一种重要的、经济有效的现代化检索手段。

三、光盘检索的作用

目前,我国CD-ROM光盘检索在情报服务和情报教育中发挥了极其重要的作用,其主要方面有:

- (1) 定题服务、回溯服务和用户检索;
- (2) 数据库套录,建立小型专题数据库;
- (3) 科研课题开题、成果申请查新和专利申请时的检索;
- (4) 普及计算机检索、机检用户的教学和培训;
- (5) 联机检索前的预检和检索方式的确定与修改。

第四节 CD-ROM光盘数据库检索提问表达式

人间是通过自然语言来沟通思想的,为了让计算机能够按照人的要求去查找材料,所以也需要一种语言来沟通人与计算机之间的“思想”,这就是计算机检索语言。光盘检索也要借助于这种计算机检索语言。这种语言同样具有一套句法,它的句法结构通常称作为“提问表达式”。任何一个提问表达式一般涉及以下三个方面。

一、提问词

例如,要查找IBM公司1982年以来计算机存储方面的文献资料,那就涉及到下列提问词:

IBM, 1982, COMPUTER, MEMORY, STORAGE, DISK, DISC, DISKETTE, FLOPPYDISK, MAGNETIC TAPE等等。

二、提问词对文献检索的要求

1. 属性要求 属性要求,又称类型要求。提问词一般有关键词、分类号、作者、年份等不同

类型,而每种光盘数据库中可作为检索项的词也有许多类型。如APS(美国专利光盘数据库)的检索项和检索词有:专利号、颁发时期、申请号、申请日期、发明人、国家地区代码、状态项、受让人、相关专利号、美国专利分类号、国际专利分类号、标题和文摘(关键词)。这样,我们就可以对前面列出的那些提问词加以归类。DIBME可以是作者,也可以是受让人;D1982E是与申请日期有关的;其它的词则可能是在标题或文摘中出现的关键词。值得注意的是,各光盘数据库由于其著录格式不同,其检索项也不尽相同,因此有必要了解光盘数据库的著录格式,以便提出有针对性的检索提问词。

2. 一致条件 对每个检索提问表达式中的提问词还必须指出与数据库的文献检索词进行比较的部位,这就是所谓的一致条件。一致条件一般分为4种:完全一致、前方一致、后方一致、任意一致。

(1) 完全一致。所谓完全一致就是指一个提问词和一个文献检索词的各个字符都必须一样。如“COMPUTER”这个词,只能查找文献中的“COMPUTER”这个文献检索词,而不能查到文献中的“COMPUTERS”,如图1-1所示。

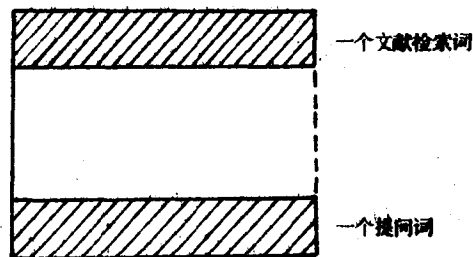


图 1-1 完全一致

(2) 前方一致。所谓前方一致就是指一个提问词与一个文献检索词相比,只要提问词与文献检索词的前面部分相一致即被认为命中。前方一致通常用通配符“*”或“?”来表示。如“SAV*”则表示“SAVE”、“SAVED”、“SAVING”、“SAVINGS”等词,而“SAV?”则只能命中“SAVE”。可见,“*”表示可通配任意个字符,而“?”则表示只能通配1个字符。如图1-2所示。

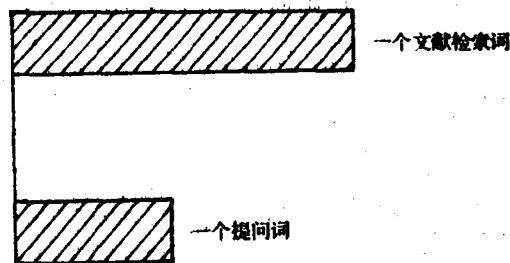


图 1-2 前方一致

(3) 后方一致。所谓后方一致就是指一个提问词和一个文献检索词相比,提问词是文献检索词的后面部分。后方一致也是用通配符“*”或“?”来表示的。如“*COMPUTER”则可表示“COMPUTER”、“MINICOMPUTER”等词,如图1-3所示。