

计算机通信 与信息检索

孙济庆 等编著

华东理工大学出版社

INFORMATION
INFORMATION
INFORMATION

546232



90546232

计算机通信与信息检索

孙济庆 葛巧珍 唐永林 编著
侯丽英 陈 荣



华东理工大学出版社

882042

(沪)新登字 208 号

计算机通信与信息检索

孙济庆 等编著

华东理工大学出版社出版发行

上海市梅陇路 130 号

邮政编码 200237 电话 64250306

新华书店上海发行所发行经销

上海展望印刷厂印刷

开本 850×1168 1/32 印张 6 字数 156 千字

1997 年 10 月第 1 版 1997 年 10 月第 1 次印刷

印数 1—5000 册

ISBN 7-5628-0819-8/TP·96 定价 9.00 元

内 容 提 要

由于计算机网络的普及与发展,利用计算机网络获取文献信息将成为人们获取信息的基本手段之一。本书全面阐述了计算机检索的手段与方法、国内外著名的计算机光盘文献信息数据库及其检索方法、联机检索、计算机通信等内容,适宜作为大专院校计算机检索课程的教材和作为科技人员信息检索的参考书籍。



清华大学出版社

前 言

为适应人们获取科技信息的需要,我们根据长期工作和教学实践,汇集了国内外的主要科技文献信息数据库的检索方法及常用的计算机检索手段,编写了《计算机通信与信息检索》一书。

本书以常用的光盘数据库的检索方法为主,突出介绍国际著名的文献信息数据库和国内常用数据库的检索方法,以及光盘网络的基本结构和目前常用的计算机检索手段——计算机通信等。本书适宜作为全国各高等院校的文献检索课程、计算机检索课程的教材,也可作为科研机构、工矿企业科技人员检索国内外科技信息的参考书籍。本书由华东理工大学多位教师合作编写。第一章、第二章由孙济庆编写,第三章由侯丽英编写,第四章由葛巧珍编写,第五章由唐永林、陈荣编写,第六章由唐永林编写,第七章由葛巧珍、侯丽英编写。全书由孙济庆审定。

由于当前信息技术发展迅猛、日新月异,加之我们的水平有限,书中难免有疏漏之处,恳切希望广大读者多提宝贵意见和建议。

编 者

1997年5月

目 录

第一章 计算机信息检索基础	(1)
第一节 计算机信息检索系统	(1)
第二节 计算机光盘数据库检索网络	(6)
第三节 计算机信息检索基本方法	(13)
第二章 计算机通信	(19)
第一节 调制解调器 (MODEM)	(19)
第二节 基于 DOS 环境的通信软件	(23)
第三节 Windows 环境下的远程通信	(33)
第三章 Dialog 光盘数据库	(53)
第一节 概述	(53)
第二节 检索指令	(58)
第三节 基本检索方法	(70)
第四节 检索实例	(75)
第四章 美国化学文摘(CA)光盘数据库	(78)
第一节 DOS 环境下的 CA 检索系统	(79)
第二节 WINDOWS 界面的 CA 检索	(91)
第五章 其他光盘数据库	(105)
第一节 《美国专利》光盘数据库	(105)
第二节 银盘信息公司数据库	(117)
第三节 ISI 公司光盘数据库	(121)
第四节 UMI 公司光盘数据库	(127)
第五节 《中文科技期刊篇名》光盘数据库	(131)
第六节 《中国经济新闻库》(XWK)	(136)
第七节 中国化学化工文献数据库(CCBD)	(137)
第六章 计算机检索的策略与技巧	(145)
一、检索时提问的表达形式	(145)
二、编制检索策略的基础与前提	(147)

第一章 计算机信息检索基础

第一节 计算机信息检索系统

一、概述

多少世纪以来,人们都是从书刊、杂志等物理介质中获得知识和信息。成千上万种图书需要组织和存放,于是就出现了人们生活中不可缺少的图书馆。这种以物理载体为基础的信息存储和传播方式一直占据着主导地位。

然而,随着计算机技术的出现与发展,人们获取信息的方式和途径也在改变着。首先出现的就是六七十年代兴起的,而现在为人们所熟知的联机信息服务。即通过远程登录对数据库进行联机检索。这种信息获取的方式 30 年来为科学的进步发挥了巨大的推动作用。然而,随着信息时代的到来,这种获取信息的方式已越来越不能适应人们对日益膨胀的信息需要。首先,远程数据库所包含的往往只有简单的文字信息,而没有文献本身;其次,昂贵而低速的通信线路也曾使许多人望而却步。

80 年代起,随着计算机存储介质的发展,文献资料的数字化,电子出版物的大量出版发行,光盘数据库的出现,使计算机信息检索摆脱了昂贵的通信线路的束缚。光盘数据库检索系统得以流行,大大方便了人们利用计算机来获取信息。由于光盘检索系统的低投入高效益,使我国光盘数据库的利用率大大高于联机检索,成为我国最主要的计算机信息检索资源,也使计算机检索系统在我国得以普及。因而本书也以光盘数据库为主介绍计算机信息检索方法。虽然光盘数据库的出现使计算机信息检索得以普及,然而仍

不能满足人们对书刊全文的需求。

近年来,随着多媒体技术,计算机网络技术和数据库技术的飞速发展,信息存取的方式面临着深刻的变化。电子存储技术的不断突破,使电子存储设备的性能价格比大幅度提高,其价格甚至已经降到比能存储同样信息的纸和胶片的价格更低,而且更安全,保存的时间更长。这就使得将大量的书刊资料存放于计算机成为现实。同时,由于价廉物美的网络服务器,宽带网络,界面友好的多媒体浏览器的出现,可将信息服务的质量提高到一个全新水平。数字化图书馆这一变革信息存取方式的时代已经到来。使人们在计算机中所获取的不仅仅是文献的信息,而是文献的本身,从而极大地丰富了计算机信息检索的内涵。“数字化图书馆”的提出与发展,使之成为矗立在信息高速公路两旁的大厦,为人类的信息利用展示了美好的前景。

二、计算机联机检索系统

计算机联机检索系统一般是以大型计算机为核心,大容量存储设备为基础,以公共数据交换网或电话通讯线路等公共通信设施为传输媒介建立起来的大型计算机情报检索系统。世界各地的用户可通过公共通信线路,用自己的微机根据系统的授权,进行远程登录,就能成为系统的仿真终端,检索系统的数据库信息。其基本模式如图 1-1 所示:

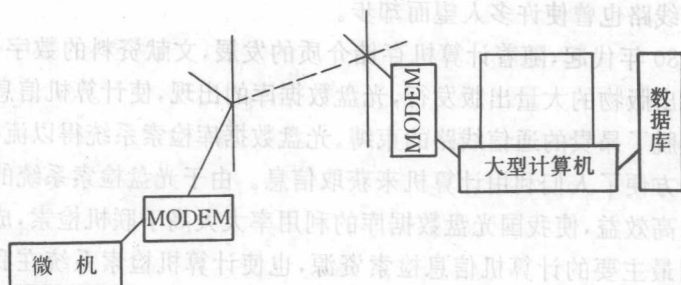


图 1-1

如图所示,在用户微机上安装计算机通信软件,经操作调制解调器呼叫系统中心的通信端,应答后,该微机即可成为系统的仿真终端进入系统的数据库检索信息。

联机检索系统起始于 60 年代,至今已有 30 余年的发展历史,是信息检索系统中最早的一种模式。联机检索系统数据容量大,数据更新快,检索方便内容丰富,而受到广大用户欢迎。

目前仍是信息检索系统的主要模式。尤其在当前国际互联网络(Internet)的全面发展情况下,在 Internet 网上有大量的这类系统可供人们检索,成为人们网上获取信息的主要来源。

由于计算机联机检索系统规模大,系统建设与维护的费用很高,这类信息检索系统大多数都需收费,而且费用比较高,目前常用的联机检索系统大部分是美国的系统,著名的有 BIALOG 系统,ORBIT 系统,OCLC 系统等等。美国建立的联机检索系统信息量大,数据库丰富,为全世界提供联网服务。在我国也有很多这类系统的终端用户。由于我国的终端用户使用这些系统需花费较高的通信费用和数据库使用费,因而联机检索的使用率较低。

联机检索系统的技术主要集中在系统中心,在用户端只要有微机,通信软件及相应的通信条件,即可联网进入系统检索信息,用户使用十分便利。

三、光盘数据库检索系统

光盘是一种存储量大,价格低廉的计算机信息存储设备,它是利用激光、计算机、数字通信和光电集成等现代高科技技术的产物。70 年代光盘被应用于音乐方面而得以迅速发展,激光唱盘目前仍是光盘的主要产品之一。80 年代中期光盘被应用于文献信息领域,具有与 IBM/PC 微机接口的光盘软件和硬件系统达到了商品化,研制成了记录文献信息的 CD-ROM 光盘(Compact Disc—Read Only Memory),使光盘进入了计算机数据库领域。目前记录有文献信息数据库的光盘在图书馆中得以广泛应用。

CD-ROM 光盘的信息存储量很高,一张普通的 5 英寸光盘,

信息存储量达到 650 余兆。相当于 3 亿多个汉字,大大高于计算机磁盘。光盘类数据存储单元,数据在制作过程中固定在其物理介质上,不能抹掉也无法修改,一般是由数据库供应商提供,并由数据库供应商用换盘的方式进行数据更新。CD-ROM 光盘是一种单面盘,其正面印刷有数据库名称与出版商版权信息等内容,反面记录着数据信息。数据信息是由物理介质上信息轨道中的一微米以下的小孔组成,当阅读器发射的激光束聚焦在信息轨道上时,在有小孔和无小孔的位置上形成不同的光反射。这两种不同的光反射经过光学系统接收后,转换成电信号,经计算机解码后,即可成为计算机的二进制原记录的数字信息。光盘由保护膜封装,不易损坏。由于阅览过程中,其盘面只与光束接触,反复阅读对盘面也无损伤,因而光盘的保存年限和读盘性能大大优于磁盘。

以光盘数据库为基础的光盘数据库检索系统,结构很简单,任何一台只要安装有光驱的 PC 机,只要放上光盘数据库,即可成为光盘数据库检索系统。光盘检索系统简单易用,投资低而得以普及。虽然光盘检索系统的出现要比联机检索系统晚十余年,但目前我国的普及率要大大高于联机检索。

近年来,光盘数据库又有了长足的发展,世界各地出版有数千种光盘数据库。数据库的规模也越来越大,一个大型数据库往往有十几张光盘,单机的光盘数据库检索系统已不能适应数据库发展的需要。90 年代开发了光盘塔和光盘网络软件,使光盘数据库检索向网络化方向发展。我国各单位建立的光盘数据库检索网络系统已逐渐占据计算机信息检索系统的主导地位,成为人们最常用信息检索系统。在下一节中将详细介绍计算机光盘数据库检索网络。

四、数字化图书馆

数字化图书馆(Digital Library)。即数字化图书信息资料库,是随着计算机网络技术、数据库技术、多媒体技术的发展应运而生的—门综合技术,它是以网络为中心的计算机时代的重要研究课题

和核心技术之一。目前,已经有许多发达国家都将其列为国家信息基础设施中的一个重要的组成部分。

1. 数字化图书馆的技术特点

1995 年 IBM 公司发起数字化图书馆的倡议,提出将信息存储、管理、查询、技术检索和发布集成在一起的立案,信息一经数字化就可以在 Internet 这样的公众网络上实现共享。方案的规模可以大小随意,以帮助所有的信息拥有者,使信息体现出最大的价值,并在全世界的网络上传播。数字化图书馆由五个方面的技术组成:数字化信息内容的制作和获取、存储与管理、检索与查询、信息的分发技术和安全和权限管理技术。

数字化信息内容的制作和获取 这是数字化图书馆的基本功能,随着多媒体技术的发展,保存存储在物理介质上的信息(如书刊)利用图、文、声、像等制作工具转化为数字化信息。

信息存储与管理 在大量的物理媒体上的信息被扫描、被照相、被识别并数字化后进行有效地存储和组织起来,以便快速地检索和查询,提供给用户简单而友好的界面,用户不必考虑信息对象的物理位置,就能从网络上便捷地访问到所需信息。

检索与查询 在数字化图书馆中,除常规的检索工具外,包含有较为高级的搜索工具。不但可对以文本形式存储的信息对象的搜索,而且提供先进的基于数字化图象,影像或声音内容的查询技术。

信息分发技术 数字化图书馆为信息拥有者提供了多种发布和传播信息的途径,可通过任何一条现有的网络、客户/服务器(Client/Server)系统,联机检索服务系统或 Internet。信息发布方式将随着用户需求的不同而变化。

信息安全和权限管理技术 为保护信息拥有者的知识产权,数字化图书馆不但提供通过网络便捷地访问信息,而且采用防止信息被非法使用的最有效的权限管理技术,这些技术包括:高级鉴定技术、版权管理、加密与水印技术,使信息产权拥有者的权益不受侵犯。

2. 数字化图书馆技术模型

数字化图书馆不是一种产品,而是一种解决方案,其信息设施随着应用客户的不同结构也可有差异,而一个典型的数字化图书馆系统基本有两个部分组成,即数字化图书馆 c/s 模式的三角形结构如图 1-2 所示:

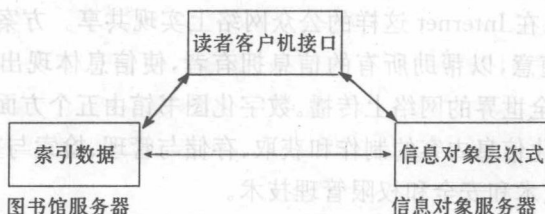


图 1-2

信息资源服务器是一个存储各类信息的“仓库”,也可以是生产各类数字化信息的工厂,其特点是具备庞大的信息存储设备,提供全文数据信息。信息存储设备一般采用 Jukebox(光盘自动存取柜),RAID(硬盘阵列)等可扩展大容量存储设备,根据用户请求,或由图书馆服务器转发的用户请求,将信息对象发送给用户。

图书馆服务器,是指客户联网的一个本地图书馆,如大学里的图书馆,它既能处理图书馆的日常事务,如:读者管理、存取控制等,而且也包含着一个数据库系统,其内容主要是文献信息的索引数据。供读者根据检索要求查询,并将读者的查询结果反馈给读者客户机,如读者需进一步索取全文,图书馆服务器就将此信息传送给信息对象服务器。

读者客户机是读者的阅读终端,通过局域网与服务器连接。在客户机上,提供了供读者进入图书馆服务器的图形界面和浏览各种信息的浏览器。数字化图书馆的客户机需对读者的身份识别,对信息的解密等功能,以确定读者获取信息的权限。

第二节 计算机光盘数据库检索网络

计算机光盘数据库检索网络是在计算机局域网(LAN)上安

装光盘塔,运行光盘数据库,即成为一种计算机信息检索系统,其特点是能同时在网络上运行几十张,甚至几百张光盘数据库,构成庞大的信息资源,信息量可达几万兆~十几万兆,同时供网络上各站点的用户共享。光盘数据库检索网络的出现一方面解决了光盘数据库的大量发展和单机阅览需频繁换盘给用户带来的不便,另一方面能供大量用户同时共享同一种数据库,充分发挥信息资源的效益,为计算机信息检索的普及发挥了很大的作用。

一、光盘数据库

计算机光盘数据库检索网络系统运行的是光盘数据库。随着CD-ROM的技术发展,光盘的种类也日益繁多,有多媒体光盘、计算机软件光盘、VCD、CD音乐光盘、计算机游戏光盘、文献信息数据库光盘等等,本书中所提及的光盘是指记录着文献信息的数据库光盘,它是计算机信息检索系统的主要信息资源。

1. 数据库的一般结构

不同的数据库,其结构也不相同,所提供的检索途径也有差异,但无论什么类型的文献信息数据库其基本特征是相同的。

记录是数据库中的基本文献单元,一个记录在数据库中往往记录着一篇文献的相关信息。记录越多,数据库的容量就越大,每一条记录在数据库中往往都有一个独有的记录号。

字段是组成记录的基本信息单元。每一个字段都描述文献的某一方面的特征。各个描述文献一部分特征的字段组合就成了记录。在数据库中为了识别每一个字段所表达的文献特征,往往都给予一个字段名。例如论文的题目字段其字段名为TI;论文的著者字段,其字段名为AU等。在一些数据库中,字段名也是具有一定的检索功能,并用字段名注释记录内容。因而上机检索前对数据库的字段名要有所了解。

顺排档是以固定的字段顺序,按文献的记录号排列的文档。在一些数据库的结构中,顺排档是文献信息的基本单位,全面记录着文献的全部信息。如果以顺排档为单位检索,检索速度比较慢。

其 倒排档是将顺排档各记录的相同字段抽取出来,再按一定顺序编制形成的文档。倒排档也称作索引,其目的是加快数据库的检索速度。在数据库中,建立的倒排档的字段越多,相应的检索效率就越高,但数据库的容量也就越大。

基本索引是指对数据库默认字段所编制的倒排档。其特征为:如输入的检索词不含有字段名,那么数据库就自动地进入基本索引中检索。大多数数据库都采用基本索引这一方法,其默认的字段也基本雷同,主要有论文题目、主题词、关键词、文摘这四个字段。但也有差异。

辅助索引是相对于基本索引的一些也建有倒排档的字段,其特征为:其输入的检索词前必须含有字段名,例如论文作者字段是一个常用的倒排档,但它是一个辅助索引,要检索 ARAI 写的文章,检索词前应加字段名 AU,如 $AU=ARAI$ 。如未加字段名就会将人名放到基本索引中去检索,造成误差。

2. 光盘数据库的运行方式

一个光盘数据库一般是由数据信息与执行程序两部分组成。不同公司出版的数据库,其结构、执行程序、运行环境、用户界面都不相同,因而光盘数据库与执行程序都需同一公司配套提供。提供方式有的采用刻写在数据库的同一张光盘上,有的另外提供安装软盘,不管采用哪一种方式光盘数据库对微机的系统环境都有一定的要求,都需执行安装程序后才能正常使用。

当光盘数据库放入光盘塔之后,执行程序一般都需安装在硬盘上,安装的过程一般只需记住此光盘数据库在光盘塔中的位置或光盘服务器对此数据库的配置。按系统的实际情况回答安装程序显示的信息,就能完成安装。网络版可直接将执行程序安装于网络的文件服务器,供共享。

当需检索数据库时,先执行光盘驱动器的驱动程序,然后运行数据库的执行程序,就能进入检索状态。如在光盘网络上,在光盘驱动程序之前,还需运行网络传输协议文件。

二、计算机光盘数据库检索网络的硬件与软件

光盘数据库网络是在计算机局域网(LAN)的基础上发展而来,除局域网上的各类微机与服务器及操作系统外,随着 90 年代的信息技术发展,开发有一系列的用于光盘网络的硬件与软件,其硬件主要是各种类型的光盘塔(CD-Tower),其软件是各种光盘组网软件,如:Opti-net,CD-net 等。

1. 光盘塔(CD-Tower)

光盘塔一般是由多个光盘驱动器叠加而成,配以接口卡,组合于局域网上,即可成为网络的组成部分。光盘塔主要依其叠加的驱动器不同而性能各异。

光盘驱动器以其盘片转速的不同,数据读取速度有差异,可分为二倍速、四倍速、八倍速等等。一般来说盘片转速越高,数据传输越快。由高倍速的驱动器叠加的光盘塔要优于低倍速的。目前常用的光盘驱动器有 IDE 接口驱动器和 SCSI 接口驱动器。用于光盘塔的驱动器一般都为 SCSI 接口驱动器,因而光盘塔的驱动器数目总是以 7 的倍数,如 7 个驱动器光盘塔,14 个驱动器光盘塔,21 个驱动器光盘塔等等。光盘塔有自己独立的供电系统,光盘塔电源的优劣也影响着光盘塔整体的性能。由于光盘塔上的驱动器采用 SCSI 接口的驱动器,SCSI 接口卡是光盘塔连接微机的接口。

2. 光盘组网软件

目前国际上开发的光盘组网软件品种较多。由于光盘数据库都是运行于 DOS 或 WINDOWS 下,因而开发的软件也应用于基于 DOS 或 WINDOWS 的局域网。我国在网络上常用的软件主要有三种:Opti-net,CD-net,SCSI Express。这三种软件有不同的版本,组网方式也有一定的差异。

Opti-net 软件是由 Online Computer Systems Inc. 出版,有两种不同的版本。Opti-net for NLN; Opti-net for Peer-to-Peer。前者是用于安装于 Novell 网的 Fileserver,由 Netware Supervisor 操作与控制。将网络上光盘驱动器对应于本地驱动器,软件

名称中的 NLM 名称为:Novell Netware Loadable Module. 后者用于独立组网。安装于 DOS, WINDOWS 等操作系统下, 就能自如地将网络上的光驱与工作站上的驱动器对应。目前 Opti-net 国内使用的主要是 4.0 版, 支持的网络传输协议有: IPX/SPX 和 NET-BIOS。新版本是 6.0 版, 支持的传输协议增加了 TCP/IP, 并能安装于 Windows NT。

CD-net 软件是由 Meridian Data Inc. 公司出版, 它有三种版本形式: CD-net for Peer-to-Peer, CD-net for NT, CD-net Plus 等。与 Opti-net 一样不同版本的软件用于不同的组网结构。CD-net 支持三种网络传输协议, TCP/IP, IPX/SPX, NET-BIOS。1995 年发布的是 6.0 版。

SCSI Express 软件是由 Micro Design International 公司出版, 它是 NLM 模块, 只能安装于 Novell Fileserver, 受 Net Supervisor 控制。SCSI Express 是将每一个 CDROM Drive 对应成 Netware 的卷 (Volume) 装入文件服务器供共享, 类似于硬盘的卷, 在 Workstation 上, 用 map 命令将相应的卷设定成网络驱动器。此软件对光盘数据库有一定的局限性。

3. 光盘网络的组网技术

由于采用的软件与硬件不同, 在局域网的基础上, 光盘网络可由多种组网方式。其结构大致有两种如图所示:

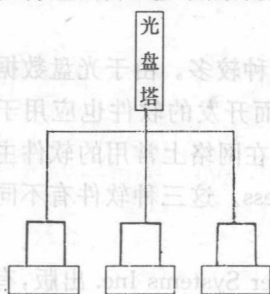


图 1-3

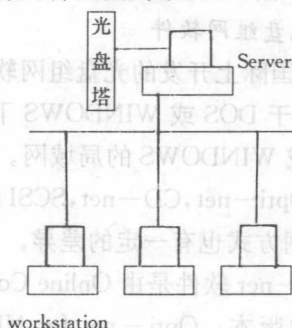


图 1-4