

电子信息资源检索

DIANZIXINXIZIYUANJIAN SUO

杨彩霞 著



中国社会科学出版社

电子信息资源检索

杨彩霞 著

中国社会科学出版社

图书在版编目(CIP)数据

电子信息资源检索/杨彩霞著.

—北京:中国社会科学出版社,2006.8

ISBN 7-5004-4641-1

I. 电… II. 杨… III. 情报检索 贾泽祥

IV. G352.2

中国版本图书馆 CIP 数据核字(2006)第 078836 号

出版发行 中国社会科学出版社

社 址 北京鼓楼西大街甲 158 号 邮 编 100720

印 刷 北京奥隆印刷厂

版 次 2006 年 8 月第 1 版 印 次 2006 年 8 月第一次印刷

开 本 787×1092 毫米 1/16

印 张 13.125

字 数 275 千字 印 数 1—2000 册 中国社会科学出版社

定 价 32 元/册

前 言

随着信息革命的迅速发展和知识经济时代的来临,信息资源已成为国家建设和企业发展的重要战略性资源,它与物质资源和能量资源一起构成现代社会经济发展的三大支柱。而电子信息资源是信息资源的重要组成部分,是数字化时代的产物。对电子信息资源的开发和利用程度是衡量一个国家经济发展水平和综合国力的重要标志之一,有效地利用电子信息资源已成为推动社会经济发展最重要的力量。在当今世界激烈的国际竞争中,谁能在电子信息资源开发利用方面占据优势,谁就能在国际竞争中夺取主动权,增强自身的竞争能力,从而促进经济的快速发展,使国家立于不败之地。

信息资源是相对于物质资源和能量资源而言的一种非物质形态的社会财富。它是依附在不同形态物质载体上的信息,经过加工和传递,能够渗入经济活动的各个方面和各个环节,对物质资源的开发和利用起着独特的作用。信息资源涉及的范围很广,表现形态各异,依附的载体也不尽相同,因此掌握和拥有电子信息资源检索的能力和程度,尤其是 Internet 的应用能力是在争夺信息资源的竞争中夺取主动权和占有优势的关键。

众所周知,Internet 是一个全球性的计算机互连网络。其中有不同学科、不同领域、不同方式、不同种类和不同性质的信息资料库。网上用户既可共享网中资源,也可将自己的信息资源发送到网上。Internet 目前已有超过 170 个国家和地区加入,直接用户超过 6000 万,成为世界上信息资源最丰富的电脑公共网络。1994 年 5 月,中国作为第 71 个国家级网加入 Internet。据 2006 年 1 月 17 日,中国互联网络信息中心(简称 CNNIC)发布的《第十七次中国互联网络发展状况统计报告》显示,截至 2005 年 12 月 31 日,我国网民人数达到 1.11 亿;宽带上网网民人数为 6430 万人,比 2004 年增加了 2150 万人,增长率为 50.2%。宽带上网成为上网接入主流。Internet 正快速地进入我国千家万户,对人们的工作、学习和生活产生着重大影响。但如何查找和检索网上数据库中的信息,如何使用网上的服务功能,对大多数人来说还比较陌生。

目前有关电子信息资源检索的书很少,所见的有关 Internet 网络的图书大多是侧重于介绍其工作原理和基本的服务功能,没有详细介绍如何查找和检索网上信息资源,因此不能满足用户的信息资源检索需求。本书针对这些问题,重点介绍各类电子信息资源的检索原理、检索策略、检索方法和检索步骤,使用户根据介绍,即可独立进行操作,独自完成电子信息资源的检索、查询、浏览、下载和打印。

由于作者的学识和水平有限,也由于网络检索手段发展的迅猛,疏漏之处在所难免,敬请同仁不吝赐教。

衷心希望,当您漫游在 Internet 这个信息高速公路上时,本书能够充当路标,帮助您一失中的,迅速在网上查找到所需的信息资料,让 Internet 真正成为每个人的信息库。

作 者
于河南科技学院
2006 年 5 月

目 录

第 1 章 电子信息资源检索概论	(1)
1.1 电子信息资源起源与发展	(1)
1.2 电子信息资源的类型与特点	(3)
1.2.1 电子信息资源的类型	(3)
1.2.2 电子信息资源的特点	(4)
1.3 数据库的类型及结构	(5)
1.3.1 数据库的类型	(5)
1.3.2 数据库的结构	(6)
1.4 电子信息资源的检索	(7)
1.4.1 数据库信息的检索	(7)
1.4.2 Internet 网络信息资源的检索	(8)
第 2 章 电子信息资源的检索方法与技术	(10)
2.1 检索方法	(10)
2.1.1 信息需求分析	(10)
2.1.2 制定检索策略	(10)
2.1.3 实施技术策略和索取原始信息	(12)
2.2 检索技术	(13)
2.2.1 布尔逻辑检索技术	(13)
2.2.2 词位检索技术	(14)
2.2.3 截词检索技术	(15)
2.2.4 限制检索技术	(15)
2.2.5 全文检索技术	(15)
第 3 章 全文数据库的检索	(16)
3.1 中文科技期刊数据库	(16)
3.1.1 概况	(16)
3.1.2 检索技术	(17)
3.1.3 检索方法和检索步骤	(18)
3.2 中国期刊全文数据库	(23)
3.2.1 概况	(23)
3.2.2 检索技术	(24)
3.2.3 检索方法和检索步骤	(24)
3.3 万方数据资源系统	(32)
3.3.1 概况	(32)
3.3.2 检索技术	(33)

3.3.3 检索方式和检索步骤	(34)
3.4 Elsevier Science	(37)
3.4.1 概况	(37)
3.4.2 检索技术	(38)
3.4.3 检索方法和检索步骤	(38)
3.5 Applied Science & Technology Plus	(44)
3.5.1 概况	(44)
3.5.2 检索技术及特点	(44)
3.5.3 检索方法	(45)
3.5.4 检索结果的显示与输出	(47)
3.6 ProQuest Digital Dissertations	(47)
3.6.1 概况	(47)
3.6.2 检索技术	(48)
3.6.3 检索方法和检索步骤	(48)
第4章 题录、文摘数据库的检索	(55)
4.1 Ei village 2	(55)
4.1.1 概况	(55)
4.1.2 检索技术	(56)
4.1.3 检索方式及检索步骤	(57)
4.2 Cambridge Scientific Abstracts	(62)
4.2.1 概况	(62)
4.2.2 检索技术	(63)
4.2.3 检索方式及检索步骤	(64)
4.3 EBSCO 联合西文期刊篇名目次库	(71)
4.3.1 概况	(71)
4.3.2 检索方法	(72)
4.3.3 检索条件输入方法说明	(73)
4.3.4 检索结果输出	(74)
4.4 BIOSIS Previews 数据库	(75)
4.4.1 概况	(75)
4.4.2 ISI 平台的 BIOSIS Previews 数据库检索方法和检索步骤	(75)
4.4.3 OVID 平台的 BP 数据库检索方法和检索步骤	(78)
4.5 OCLC First Search 数据库	(83)
4.5.1 OCLC 概况	(83)
4.5.2 检索方法和检索步骤	(84)
第5章 引文索引信息检索	(92)
5.1 中国科学引文数据库(CSCD)	(92)
5.1.1 概况	(92)
5.1.2 检索方式和检索步骤	(94)

15.2 中文社会科学引文索引(CSSCI)	(101)
15.2.1 概况	(101)
15.2.2 检索方法	(102)
15.3 ISI Web of Science(SCI、SSCI、A&HCI)	(104)
15.3.1 美国 ISI 概况	(104)
15.3.2 检索技术	(107)
15.3.3 检索方式及检索步骤	(108)
15.3.4 其它检索功能	(116)
15.4 美国《科技会议录索引》(ISTP)	(119)
15.4.1 ISTP 概况	(119)
15.4.2 《科技会议录索引》光盘检索	(119)
15.4.3 《科技会议录索引》网络版检索方法与检索步骤	(122)
第 6 章 专利信息检索	(126)
6.1 中国专利	(126)
6.1.1 国家知识产权局专利数据库	(126)
6.1.1.1 概况	(126)
6.1.1.2 检索技术	(126)
6.1.1.3 检索方法及检索步骤	(126)
6.1.2 中国专利信息网	(127)
6.1.2.1 概况	(127)
6.1.2.2 检索方式	(127)
6.2 美国专利	(129)
6.2.1 概况	(129)
6.2.2 美国专利与商标局专利数据库	(130)
6.2.2.1 检索方法及检索步骤	(130)
6.2.2.2 检索字段	(132)
6.3 欧洲专利	(133)
6.3.1 概况	(133)
6.3.2 欧洲专利局数据库	(133)
6.3.2.1 检索方式及检索步骤	(134)
6.3.2.2 检索实例	(136)
6.4 日本专利	(138)
6.4.1 概况	(138)
6.4.2 工业产权数字图书馆(IPDL)	(139)
6.4.2.1 检索方式及检索步骤	(139)
6.4.2.2 检索实例	(141)
6.5 加拿大专利	(141)
6.5.1 概况	(141)
6.5.2 加拿大专利数据库	(141)

6.5.2.1 检索方式及检索步骤	(141)
第 7 章 电子图书及其检索	(144)
7.1 电子图书概述	(144)
7.1.1 电子图书的概念及发展	(144)
7.1.2 电子图书的优势	(144)
7.1.3 电子图书的类型	(145)
7.2 超星数字图书馆	(146)
7.2.1 概述	(146)
7.2.2 超星电子图书的内容	(146)
7.2.3 超星电子图书系统的功能	(147)
7.2.4 超星电子图书的使用	(148)
7.3 书生数字图书馆	(151)
7.3.1 概述	(151)
7.3.2 书生电子图书的内容	(152)
7.3.3 书生电子图书系统的功能	(152)
7.3.4 书生电子图书的使用	(152)
7.4 方正 Apabi 数字图书馆	(154)
7.4.1 概述	(154)
7.4.2 Apabi 电子图书的内容	(155)
7.4.3 Apabi 电子图书的功能	(156)
7.4.4 Apabi 电子图书的使用	(156)
7.5 免费电子图书	(157)
第 8 章 电子期刊及其检索	(161)
8.1 电子期刊的产生和发展	(161)
8.2 电子期刊的特点与类型	(161)
8.2.1 电子期刊的特点	(161)
8.2.2 电子期刊的类型	(164)
8.3 电子期刊选介	(165)
8.3.1 常用中文电子期刊	(165)
8.3.1.1 中文科技期刊数据库	(165)
8.3.1.2 中国期刊全文数据库	(165)
8.3.1.3 万方数据资源系统中的数字化期刊子系统	(166)
8.3.1.4 《人大复印报刊资料》全文数据库	(166)
8.3.2 常用外文电子期刊	(166)
8.3.2.1 Nature	(166)
8.3.2.2 Science	(168)
8.3.2.3 Academic Press 的电子期刊	(172)
8.3.2.4 Elsevier Science 电子期刊	(174)
8.3.2.5 Springer Link 全文电子期刊	(175)

8.3.2.6 IEEE/IEE 电子期刊	(183)
8.3.2.7 英国 Maney 公司和英国皇家学会网络版期刊网	(186)
8.3.2.8 英国皇家物理学会的电子期刊	(186)
第 9 章 电子报纸及检索	(188)
9.1 电子报纸的概况	(188)
9.2 国内电子报纸的检索	(188)
9.2.1 书生之家	(188)
9.2.2 人民网	(189)
9.3 国外电子报纸的检索	(192)
9.3.1 国外电子报纸的检索	(193)
9.3.2 国外电子报纸的阅读	(193)
9.4 电子报纸的发展	(194)
9.5 免费电子报纸	(195)
参考文献	(199)

第1章 电子信息资源检索概论

电子信息资源是信息资源的重要组成部分,是电子化了的信息资源,它是以数字化的形式,把文字、图形、图像、声音、动画等多种形式的信息存放在光、磁等非印刷型介质上,以电信号、光信号的形式传输,并通过计算机、通讯设备及其他外部设备再现出来的一种信息资源。

电子文献是信息社会文献载体形成、发展、进化的必然产物。在整个社会的文献消费体系中,电子文献消费指数正在成倍地增长,并逐步动摇着纸质文献长期以来在文献消费体系中的主导地位。电子信息产品更是琳琅满目、丰富多彩。从结构上看,既有学术性、商业性内容,也有文学性、教育性和娱乐性内容;既有书目型、全文型产品,也有阅读型、视听结合型和多媒体对话型产品;既有正式出版物,也有个人信函资料、档案纪录等非出版物。有精品,也有粗制滥造的平庸之作,更不乏低级趣味的内容,可谓林林总总。

电子信息资源自1946年世界上第一台大型电子计算机ENIAC诞生之后,便得以从信息资源中脱颖而出,并在现代通信技术、计算机技术的飞速发展的情况下得到蓬勃发展的。随着通信技术、多媒体技术和计算机技术的迅猛发展,特别是Internet的逐渐成熟和网络资源大量涌现,Internet已成为人们进行科学研究、商业活动和共享信息的重要手段。由于Internet上的所有主机都采用TCP/IP协议连接和通信,网络的各种计算机也都遵循该协议规定的方式进行数据交换,因而互联网成为各种数据库的总数据库,可以称得上是世界最大的电子信息资源库。

1.1 电子信息资源起源与发展

ENIAC采用0和1来表示数据和指令,从而拉开了数据信息电子化的序幕。但最初计算机的功能也只是进行数据的计算,随着计算机技术的不断进步,计算机才逐渐开始用于信息的处理,而信息只有积累到一定的量才能称之为资源,所以我们说,最初计算机进行处理的并不是电子信息资源。电子信息资源的真正起源是美国国立医学图书馆在上世纪60年代中期用电子计算机出版的《MEDLARS》(Medical Literature Analysis and Retrived System)数据库,它被用于医学文献的检索。随后,美国其他一些机构也相继使用电子计算机编辑文摘期刊从而在客观上编制成了机器可读的书目文档,称之为文献目录型数据库。这种文献目录型数据库由于使用计算机来编辑、检索文献信息,因此我们可以将机编文献目录型数据库的研制视为电子信息资源形成的开端。(机编文献目录数据库信息阶段)

随着科技发展的需要,这种目录型数据库的专业面越来越广,除自然科学、社会科学以外,还包括经济及人文科学等。另外,还出现了与目录型数据库类似的存取数值型信息的数值型数据库和全文记载事实的事实型数据库等。但这些数据库只能进行批量检索,即检索服务部门把许多用户的课题及其要求加以汇总,进行批量检索,然后把检索结果邮寄给检索用户。显然,其缺点是用户不能进行人机对话及时修改检索策略,更不能及时获得检索结果。通信网络的出现和发展则弥补了这一大缺陷,它促进了电子信息资源远距离的传输和交换。首先是银

行界、建筑行业以及商业界为了得到准确及时的行情和金融方面的信息,率先使用了计算机处理技术,并进一步推动了公用通讯网络与计算机系统相结合的国际联机检索系统。国际联机检索就是用户使用终端设备,远距离地从国际联机检索中心迅速而准确地获取电子文献信息,使知识或信息得到广泛而有效的传播和利用,其实质就是数据库和通信的结合。从上世纪六七十年代起,许多国家还先后建立了专门从事计算机检索的机构,如美国的洛克希德公司(Lockheed)和系统发展公司(System Development Corporation),英国的目录检索服务处(Bibliographic Retrieval Service)和图书馆自动化情报服务处(British Library Automated Information Service)与 Infoline 公司以及欧洲的 DIMDI(Deutsches Institut für Medizinische Dokumentation und Information)和 ESA(European Space Agency)等,这些机构都建有大量的数据库联机检索系统,都向全世界联机用户提供电子信息服务。其中,著名的系统有 Dialog、ORBIT、ESA—IRS、JOIS、BRS、QUESTEL、OCLC、STN 等。以 Dialog 系统为例,美国洛克希德飞机与导弹公司的 Dialog 系统是目前美国和世界上最大的检索系统。Dialog 系统于 1966 年正式开展文献检索,当时其数据库存储的文献只有 20 多万篇。随着时间的推移,其数据库的数量在逐年增加,数据库范围涉及到自然科学和社会科学的近 20 个学科,50 多种语言。文献类型有:学位论文、会议录、科技报告、政府文件、专利标准、经济预测、公司行业名录、私人文档和统计数据等。由于联机服务和数据库产品迅猛发展,使电子信息资源引起了人们的普遍关注。但国际联机检索系统能够传输、检索的仍主要是电子文献信息资源,还停留在电子文献信息阶段。(国际联机检索信息阶段)

利用国际联机检索系统检索到的电子文献信息具有较高的实用价值。但国际联机检索费用昂贵,一般用户难以负担,人们开始努力寻求一种低廉的存储、检索电子信息的方式,光盘存储技术则适应了这一需求。光盘是上世纪 70 年代出现的一种新型存储设备。近十年来,由于许多尖端技术领域所取得的新成就,特别是光学、光电子技术、微电子技术的发展,光盘技术也得到了飞速的发展。由于光盘在存储电子信息资源方面具有记录密度高、容量大、成本低、体积小、寿命长、可实现随机存取和检索费用低廉等优点,因此,光盘被广泛用于存储、检索电子信息资源,并产生了一批生产系列光盘的公司,如美国 UMI 公司和银盘公司等。光盘记载的电子信息资源并不局限于文献信息,还包括各种软件,但可用于检索的仍以文献信息为主。(光盘信息阶段)

国际联机检索和光盘检索为我们提供了大量的电子信息资源,但各自又都有着或多或少的缺点,例如国际联机检索的费用昂贵,检索指令复杂,多为专业检索人员才会使用,而光盘检索得到的信息又不是十分及时等。因此,人们渴望一种新型检索方式的出现。1993 年,美国政府提出 NII 计划(国家基础设施建设),兴建以 Internet 为雏形的“信息高速公路”。随着 Internet 的迅猛发展,网络资源如潮水般涌来,它已成为人类进行科学研究、商业活动和共享信息资源的重要手段,并为世界各国所接受。Internet 是网络的网络,是网络通过互联而形成的全球网。它已延伸到地球上几乎每个国家。在 Internet 网上的所有主机都采用 TCP/IP 协议连接和通信,使网上各种计算机都遵循该协议所规定的方式进行数据交换,其结果是使得 Internet 有极为丰富的电子信息资源,堪称世界上最大的图书馆或信息资源库。Internet 信息资源主要包括电子报刊、电子新闻、电子报告、电子论坛、会议资料、各种软件资料、图像文件、声音文件和电子游戏等。由于 Internet 起源于美国军方,并在学术界得到了迅速发展,美国政府扶持的力度较大,因此网中大部分资源是免费的,如 Archie 数据库、WAIS 数据库、大多数大

学图书馆的书刊数据库以及 WEB 信息系统中的共享性数据库等,可以说,使用 Internet 所提供的电子信息资源费用较为低廉。另外,Internet 的信息更新速度快,其信息在随时刷新,用户得到的信息总是最新的。此外,Internet 中的 WWW 网络信息系统采用链式结点链接,用户检索起来不但十分方便,而且检索范围大大拓宽了,可检索的信息也不再以文献信息为主。同时,Internet 还提供了多种信息服务,主要有 E-mail、Telnet、WWW 和 BBS 等。为便于充分利用丰富的信息资源,Internet 上还提供各种各样的检索工具,如 Archie 数据库检索系统和 Gopher 菜单式数据库检索系统等。总之,Internet 是目前世界上资料最多、门类最全、规模最大的信息库,是人们获取信息的重要来源。(网络信息阶段)

1.2 电子信息资源的类型与特点

1.2.1 电子信息资源的类型

从信息检索的角度,电子信息资源主要是指通过计算机等设备以数字信号传递的数字信息资源。数字信息资源按不同的划分标准,可分为下列几类。

一、按信息的载体和通讯方式分

(1)联机信息资源。由计算机联机信息服务系统提供的信息资源,如著名的 Dialog、STN 和 ORBIT 系统,为全世界用户提供了丰富的电子信息资源。用户使用检索终端设备,通过通信设施(如通信网、调制解调器、自动呼叫器、通信控制器等),直接与中央计算机相连,检索远程数据库中的信息资源。在检索过程中是人—机对话式,可随机或脱机浏览、传递所得信息,并可根据需求随机修改检索策略,具有实时、快速、信息追溯年代长、查准率高的特点,但检索费用较昂贵。

(2)光盘信息资源。光盘信息资源主要包括各种信息数据库,它有单机版可进行单机检索,有网络版可在网上进行检索,还可以与联机检索系统联网进行联机检索。光盘信息资源的检索过程与联机检索极为相似,主要不同之处是信息追溯年代不如联机系统检索年代长,但检索费用大大低于联机费用,在通信不发达的地区,利用单机光盘信息资源是一种获取信息的有效途径。

(3)Internet 网络信息资源。Internet 是通过标准通信方式(TCP/IP 协议)将世界各地的计算机和计算机网络互联而构成的一个结构松散、交互式的巨型网络。它目前是世界上最大的信息资源库,其中大部分资源是免费的,与前两类相比,具有信息更新速度快,随时都在刷新的特点,用户得到的信息总是新的。此外,Internet 中的万维网—WWW 信息系统(又称 Web 信息系统),采用客户机/服务器的模式,以超文本的方式链接分散在 Internet 上 WWW 服务器中的信息。用户通过 WWW 浏览器可以方便地访问网上遍布世界各地的 WWW 服务器中的信息。WWW 信息服务的特点是在网上进行多媒体信息的收集、分类、存放、发布和交流,并向网上的用户提供信息检索及其它交互式服务,它与传统的网络信息服务的区别在于:从提供信息的形式上,WWW 服务器上提供的是多媒体信息,在页面设计上具有结构合理、可读性强、用户界面友好程度高等优点,而传统网络信息服务提供的信息是单一的。从信息内容上看,WWW 服务器上提供的信息是包罗万象的,而传统网络信息服务提供的信息大部分是行业性的。从提供信息服务的情况看,WWW 服务器上提供的信息更具有及时性,入网费用也低于传统的网络信息服务。在网页上展现的各种链接可随时让用户获得更多的相关信息。

二、按信息的表现形式分

(1) 文本信息资源。一般的文本信息资源是按知识单元的线形顺序排列, 具有较大的局限性和片面性。

(2) 超文本信息资源。超文本信息资源是按知识单元及关系建立的知识结构网络。它通过网络上各节点的链路把相关信息有机地结合在一起。检索超文本信息资源时, 可从任何一个节点开始, 以知识片段及其关系作为检索、追踪的依据。

(3) 多媒体信息资源。是包括文本、图像和声音等各种信息表达的总称。多媒体信息资源能使人们获得的信息不仅同具图、文、声, 而且还丰富多采。

(4) 超媒体信息资源。是超文本和多媒体技术的结合, 具有超文本和多媒体两种信息资源的特点, 具有高度的交互性。在超媒体信息系统中, 不同类型的媒体信息能高度综合和集成, 空间上图、文、声并茂, 时间上媒体信息同步实现。

1.2.2 电子信息资源的特点

一、是数字化的信息资源

信息资源由纸张上的文字变成磁性介质上的电磁信号或光介质上的光信息, 使信息的存储传递和查询更加方便, 而且所存储的信息密度高, 容量大, 可以无损耗地被重复使用。以数字化形式存在的信息, 既可以在计算机内被高速处理, 又可以通过通信网络进行远距离传送, 这就使全球信息资源共享成为可能。随着网络的进一步扩大与应用范围的拓宽, 数字化的信息资源将成为信息资源的最终转化方式。

二、以网络为传播媒介

传统的信息存储载体为纸张、磁盘、磁带, 而在网络时代, 信息的存在是以网络为载体, 人们得到的是网络上的信息, 而不必过问信息是存储在磁盘上的还是磁带上的。

三、内容丰富多样

从过去单一的文字信息, 到今天图形、图像、声音等多种信息, 涉及的领域也非常广阔, 从经济、科研、教育、艺术到具体的行业和个体信息。包含的文献类型也多种多样, 从电子报刊、电子工具书、商业信息、新闻报道、书目数据库、文献信息索引到统计数据、图表、电子地图等等。这就使得信息的组织方式发生了巨大的变化, 不仅以知识和信息为基本单元, 而且充分展示这些单元间的逻辑关系, 为网络环境下不同形式的信息资源的管理和开发利用提出新的要求。

四、信息资源数量巨大, 影响面广

据统计, Internet 网上每天发布的新信息有 15 万件, 按计算机储存字节算有 480MB, 网络信息总量在 20TB 以上。

五、数据结构具有通用性、开放性和标准化的特点

在网络环境下, 数据可以被多人同时访问, 是一种共享性的信息资源, 这就使得数字信息资源更易于实现资源的扩充。

六、传播方式的动态性

网络环境下, 信息传递和反馈快速灵敏, 具有动态性的实时性的特点。信息在网络的流动非常迅速, 电子流取代了纸张和邮政的物流, 加上无线电和卫星通信技术的充分运用, 任何信息源, 只要上传到网上, 就能够在短短的数秒内传递到世界的每一个角落。

七、信息源复杂, 灰色信息较多

书。由于网络的共享性与开放性,人人都可以在互联网上索取和存放信息,这些信息没有经过严格编辑和整理,良莠不齐,各种不良和无用的信息大量充斥在网络之上,形成了一个纷繁复杂的信息世界。这就要求信息管理机构建立一套科学高效的信息过滤体系,区分出有价值的信息加以利用。

1.3 数据库的类型及结构

1.3.1 数据库的类型

数据库是在计算机存储设备上按一定方式,合理组织并存储的相互有关联的数据的集合,是计算机技术和信息检索技术相结合的产物,是电子信息资源的主体,是信息检索系统的核心部分之一。按照国际上通用的分类方法,数据库通常划分为以下三大类:

一、参考数据库——指引用户到另一信息源以获得完整的原始信息的一类数据库。它包括书目数据库和指南数据库。

(1)书目数据库——存储描述如目录、题录、文摘等书目线索的数据库,又称二次文献信息数据库。如各种图书馆目录数据库、题录数据库和文摘数据库等属于此类,它的作用是指出了获取原始信息的线索。图书馆目录数据库,又称机读目录,其数据内容详细,除描述标题、作者、出版项等书目信息外,还提供索取原始信息的馆藏信息。题录、文摘数据库描述的数据内容与印刷型的题录、文摘相似,它提供了论文信息或专利信息等确定的信息来源。

(2)指南数据库——存储描述关于机构、人物、产品、活动等对象的数据库。与其它数据库相比,指南数据库提供的不仅仅是有关信息,还包括各种类型的实体,多采用名称进行检索。如存储生产与经营活动信息的机构名录数据库、存储人物信息的人物传记数据库、存储产品或商品信息的产品指南数据库、存储基金信息的基金数据库等属于此类,它的作用指引用户从其它有关信息源获取更详细的信息。

二、源数据库——主要存储全文、数值、结构式等信息,能直接提供原始信息或具体数据,用户不必再转查其它信息源的数据库。它可以划分为以下几种类型:

(1)全文数据库。存储原始信息全文或主要部分的一种源数据库,简称全文库。如期刊全文数据库、专利全文数据库、百科全书全文数据库。用户使用某一词汇或短语,便可直接检索出含有该词汇或短语的原始信息的全文。从信息源类型来看,它又属于文献信息库的范畴。

(2)数值数据库——存储以数值表示信息为主的一种源数据库,和它类似的有文本—数值数据库。与书目数据库比较,数值数据库是对信息进行深加工的产物,可以直接提供所需的数据信息。如各种统计数据库、科学技术数据库等。数值数据库除了一般的检索功能外,还具有准确数据运算功能、数据分析功能、图形处理功能及对检索输出的数据进行排序和重新组织等方面的功能。

(3)术语数据库——专门存储名词术语、词语信息以及术语工作和语言规范工作成果的一种源数据库,各种电子化辞书也包括在内。

(4)图像数据库——用来存储各种图像或图形信息及有关文字说明资料的一种源数据库,主要应用于建筑、设计、广告、产品目录、商标图形、图片或照片等资料类型的计算机存储与检索。

三、混合型数据库——能同时存储多种不同类型数据的数据库。

(1)多媒体数据库——把文字、数值、声音、图像等性质不同的信息存储于同个媒体上,并进行一体化处理和管理的一种新型数据库。

(2)超媒体数据库——通过外部树形的链将多种类型的媒体连成一个集合,该集合称为超媒体数据库。

(3)超文本数据库——超文本数据库类似于一般数据库而又优于一般数据库。超文本的最大特点在于用户界面是以窗口形式出现,窗口中给以图标等图形标记,用户可以通过鼠标器激活图标中的相应结点,以任意顺序通过链路阅读所有相关的资料。超文本数据库目前被普遍用于国际互联网和公共场所信息指南检索等。

1.3.2 数据库的结构

一、书目数据库的结构

书目数据库是以文档形式组织一系列数据,这些数据被称为记录,一个记录又包含若干字段。

(1) 文档(File)

文档是按一定结构组织的相关记录的集合。文档是书目数据库数据组织的基本形式,文档的组织方式与检索系统的硬件和软件功能密切相关。在书目数据库中,文档结构主要分为顺排文档和倒排文档。

● 顺排文档。记录按顺序存放,记录之间的逻辑顺序与物理顺序是一致的,相当于印刷型工具中文摘的排列顺序,是一种线形文档。顺排文档是构成数据库的主体部分,但其主题词等特征的标识呈无序状态,直接检索时,必须以完整的记录作为检索单元,从头至尾查询,检索时间长,实用性较差。

● 倒排文档。将顺排文档中各个记录中含有主题性质的字段(如主题词字段、标题字段、叙词字段等)和非主题性质字段(如作者字段、机构字段、来源字段等)分别提取出来,按某种顺序重新组织得到的一种文档。具有主题性质的倒排档,称基本索引档,非主题性质的倒排档,称辅助索引档。

综上所述,顺排档和倒排档的主要区别是:顺排档以完整的记录为处理和检索单元,是主文档,倒排档以记录中的字段为处理和检索单元,是索引文档。计算机进行检索时,先进入倒排档查找有关信息的存取号,然后再进入顺排档按存取号查找记录。

(2) 记录(Record)

记录是作为一个单位来处理有关数据的集合,是组成文档的基本数据单位。记录中所包含的若干字段,则是组成记录的基本数据单位。在书目数据库中,一个记录相当于一条题录或文摘。

(3) 字段(Field)

字段是文献记录的基本单元。一条记录有若干字段,一个字段有时还可分为几个子字段(subfield)。在书目数据库中,一条记录通常由标题字段、作者字段、来源字段、文摘字段、主题词字段、分类号字段、语种字段等组成。数据库的字段可分为基本字段和辅助字段。基本字段主要是描述文献内容特征的字段,如篇名、文摘、叙词、自由表因词等字段,辅助字段主要是描述文献外部特征的字段,如著者、机构名称、语种、文献来源等字段。

为了便于计算机识别,每个字段都有一个相应的标识符,如 TI、DE 分别表示篇名和叙词字段。在这条记录中,有的字段又由若干个子字段组成,这些子字段彼此是同等关系,在内容

上有一定的联系,但相互独立,例如,叙词字段有若干个叙词。记录的字段标识及名称说明见表 1-1。

表 1-1 数据库字段标识及名称

字段名	标识符	全称	字段名	标识符	全称	字段名	标识符	全称
文献题名	TI	Title	著者	AU	Author	著者单位	AF	Author affiliation
摘要	AB	Abstract	出版者	PB	Publisher	出版国	CP	Country of publication
叙词	DE	Descriptor	文献来源	SO	Source	出版年	PY	Publication year
关键词	ID	Identifier	语种	LA	Language	分类号	CC	Classification code
文献类型	DT	Document type	国际标准刊号	IS	ISSN	记录存储号	AN	Accession number

二、全文数据库的结构

一般的全文数据库结构与书目数据库相似,全文数据库的一个记录就是一个全文文本,记录分成若干字段。其主文档是以顺排形式组织的文本文档,倒排档是对应于记录可检字段的索引文档。

三、数值数据库的结构

数值数据库的结构要综合考虑数据库的内容及检索目的,即在内容上,数值数据库的主要内容是数值信息,但不排除含有必要的说明性的文本信息,在检索上,便于单项检索和综合检索,还能对数值进行准确的数据运算、数据分析、图形处理及对检索输出的数据进行排序和重新组织。数值数据库的数据结构可以是单元式,也可以是表册形式。前者是对原始数据的模拟,后者则是对统计表格的机读模拟。数值数据库通常有多种文档,如顺排档、倒排档、索引文档等。顺排档是由数值数据组成,为主文档,另有相应的索引文档,为便于存取,索引文档采用基本直接存取结构的组织形式。倒排档也有相应的索引文档,索引文档采取分级组织形式。数值数据库的文档结构,使所有文档都可以用于检索,所有数据都可用来运算,构成了数值数据库的特点。

四、指南数据库的结构

指南数据库的结构兼有书目数据库、全文数据库和数值数据库的特点,有顺排档、倒排档、索引文档和数据字典。一般而言,对涉及主题领域较多,内容综合性较强的大型指南数据库,顺排档(主文档)可采用多子文档的结构,对单一主题领域和内容较专的,则采用单一主文档和不定长、多字段的记录格式为宜。

1.4 电子信息资源的检索

电子信息资源的检索是指使用相关的检索技术,通过检索系统在数据库或网络信息资源中自动查找用户所需有关信息。

1.4.1 数据库信息的检索

数据库信息检索系统包括联机数据库、光盘数据库和网络数据库,检索的基本流程如图 1-4-1 所示。

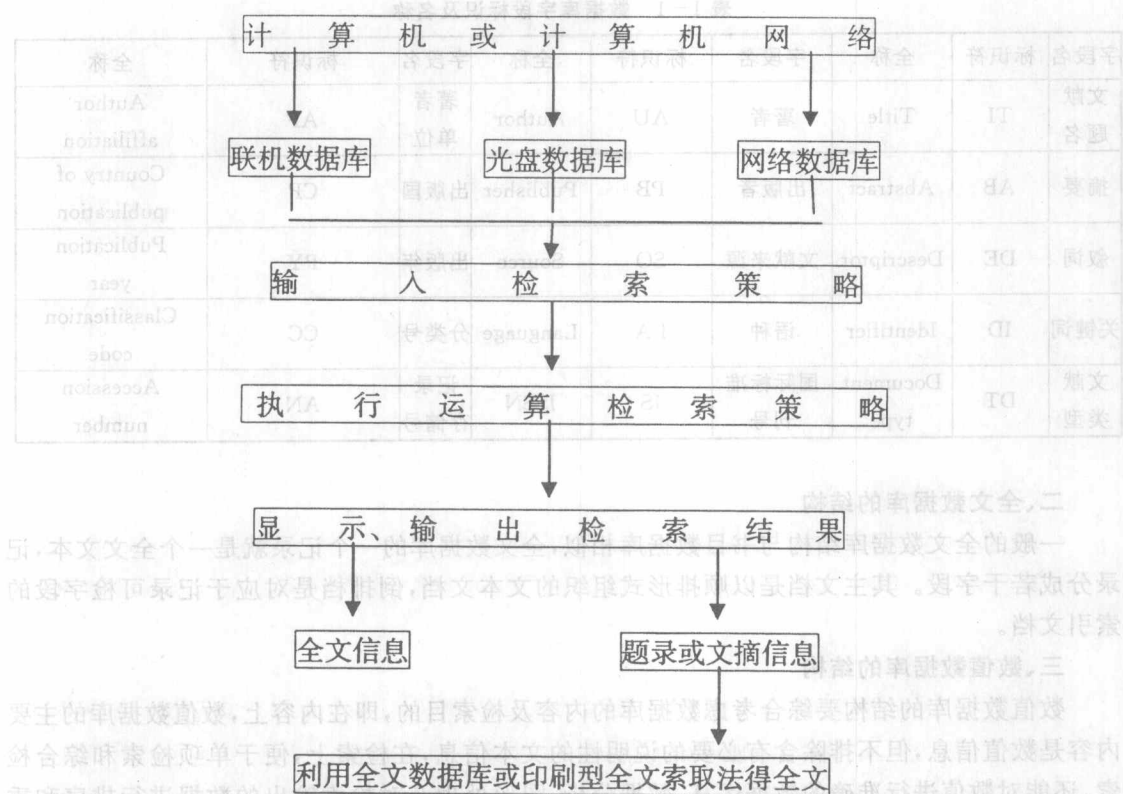


图 1-4-1 数据库信息检索基本流程

1.4.2 Internet 网络信息资源的检索

目前,在 Internet 运行的交互式信息服务软件主要基于 WWW。WWW(World Wide Web)是 Internet 上最先进的网络信息资源检索系统,它采用超文本技术、网络技术和多媒体技术,将不同类型的文件通过关键词建立链接,为用户提供一种交互式信息检索接口,具有 Internet 上现有的所有服务功能,即 Telnet、FTP、WAIS 和 E-mail 等。因此,WWW 是 Internet 上最常用的一个多媒体信息浏览与检索系统,它把 Internet 上所有的信息资源组成一系列的超文本和超媒体文件,用户可进行远程信息浏览与检索,获得静态和动态的知识信息。WWW 具有的主要特点是信息检索空间宽,即 WWW 信息服务器采用超文本技术和统一资源定位器技术,将全世界联机信息资源联接在一起,形成人类知识的结合体。它根据需要,指引进一步检索存放在其它 WWW 服务器上的相关信息。

一、WWW 网络地址和浏览器

使用 WWW 检索信息一般使用 WWW 网络地址,WWW 网络地址称为 URL (Uniform Resource Locator),即统一资源定位器,它对 WWW 服务器的信息资源,都规定一个相应的特有地址。URL 地址由五部分组成:Protocol://host.[:port]/path/filename。Protocol 是通讯协议,目前最常用的是超文本传输协议,即 Hyper Text Transfer Protocol(HTTP);host 是主