

电子信息源 与网络检索

谢新洲 编著

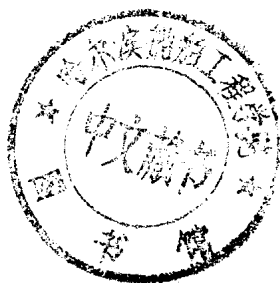
北京图书馆出版社

459022

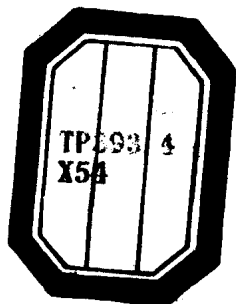
国家自然科学基金资助项目成果

电子信息源与网络检索

谢新洲 编著



5



00459022

北京图书馆出版社

图书在版编目(CIP)数据

JS208/61

电子信息源与网络检索/谢新洲编著. —北京:北京图书馆出版社,1999.1

ISBN 7-5013-1618-X

I. 电… II. 谢… III. ① 情报检索-检索方法 ② 因特网-信息-检索方法 IV. G354.2

中国版本图书馆 CIP 数据核字(98)第 40314 号

书名 电子信息源与网络检索

著者 谢新洲 编著

出版 北京图书馆出版社(原书目文献出版社)
发行 (100034 北京西城区文津街 7 号)

经销 新华书店

印刷 北京飞达印刷厂

开本 850×1168 毫米 1/32

印张 11.625

字数 290(千字)

版次 1998 年 12 月第 1 版 1998 年 12 月第 1 次印刷

印数 1—4000

书号 ISBN 7-5013-1618-X/G · 426

定价 17.00 元

序

世界有两类资源,一类是客观资源,另一类是主观资源。客观资源包括物质、能量、信息,它们各有自己的特性,如人们常说的,物质不灭,能量守恒,信息非对称。在这三种资源中,信息的相对重要性随着社会进步而不断提高。主观资源就是人本身,这是最重要的资源。从客观上看,人这类主观资源就是物质、能量、信息的特殊的统一体,它的特性是通过利用客观资源来开发自己,反过来又提高对客观资源的利用效率。世界的历史就是主观资源与客观资源相结合,各种资源开发和利用的优化过程。

数字化、多媒体的因特网(Internet)的发展和普及,把信息资源的开发和利用推进到一个崭新的阶段即网络化电子信息阶段。在这个阶段,信息达到空前丰富的程度,以至出现了从未有过的信息“爆炸”。据统计,近十年来,全球信息总量每过一年就会新增790亿条信息。^① 在信息的茫茫沧海中,如何找出宝贵一粟的所需信息,成了十分棘手的问题。怎样使现有信息发挥更大的作用,也是一个重要而又困难的问题。这要求人们学习和掌握电子信息源的有关知识,学习和掌握网络检索的相应技术。

谢新洲同志编著的《电子信息源与网络检索》,为我们提供了满足上述需求的一本实用的著书。该书既对基于因特网的电子信息源特别是数据库和网络信息作了论述,又对网络检索的各种基本技术和在因特网上搜索、浏览信息的方法、工具作了介绍,还全面、系统地说明了电子图书、电子期刊、电子报纸的内容、类型和特

^① 参见《中国民航报》1998年1月12日。

点,以及检索这类网络空间内书报刊的方法。此外,书内专辟一章,详细介绍了因特网上的国内外主要的著名图书馆。

本世纪 50 年代初,记得我分配到中国科学院经济研究所开始从事经济研究工作时,第一项训练,就是进图书馆学习如何查找书报刊的目录和索引。在 90 年代末的今天,我们要学习,要研究,就必须首先掌握(最好是熟练地掌握)在网络环境下查找电子信息的自我服务技能。如果你真想这么做,那么请你读一读这本书,相信对你会有不小的帮助。

乌家培

目 录

第一章 绪论	(1)
第二章 电子信息源	(4)
2.1 数据库	(4)
2.1.1 美国的数据库业	(5)
2.1.2 西欧的数据库业	(12)
2.1.3 中国的数据库业	(15)
2.1.4 数据库的新发展	(18)
2.2 网络电子信息源.....	(19)
2.2.1 网络电子信息源的特点	(19)
2.2.2 网络电子信息源的类型	(21)
第三章 Internet 环境下的联机检索	(23)
3.1 联机检索与 Internet 网络检索:比较分析	(23)
3.1.1 两种不同的系统结构:集中式与分布式	(23)
3.1.2 两种不同的服务模式:主仆式与客户/服务器式	(26)
3.1.3 两种不同的信息组织方式:普通文本与超文本	(28)
3.1.4 两种不同的检索机制	(29)
3.2 寻找新生长点的联机检索.....	(32)
3.2.1 联机检索系统的新发展	(32)
3.2.2 美国联机服务公司(AOL)	(34)
3.2.3 CompuServe	(42)
第四章 网络检索的基本技术	(52)
4.1 Internet 的基本状况	(52)
4.1.1 Internet 的发展历史.....	(52)

4.1.2	Internet 在中国	(57)
4.1.3	下一代 Internet 网	(60)
4.2	Internet 的基本知识	(61)
4.2.1	Internet 的概念	(61)
4.2.2	Internet 实现互连的核心——TCP/IP	(63)
4.2.3	Internet 的域名与地址	(70)
4.2.4	Internet 的连接	(74)
4.2.5	Internet 提供的主要服务	(76)
4.3	菜单式信息检索工具——Gopher	(85)
4.3.1	Gopher 的基本概念	(85)
4.3.2	Gopher 的使用	(86)
4.4	超文本信息检索工具——WWW	(89)
4.4.1	WWW 的特性与功能	(91)
4.4.2	WWW 的特点	(91)
4.4.3	WWW 的系统结构	(92)
4.5	网景公司的浏览器 Netscape	(95)
4.5.1	Netscape 的使用	(95)
4.5.2	Netscape 常用下拉菜单的功能	(97)
4.5.3	Netscape 常用键盘命令	(102)
4.6	微软公司的浏览器 Explorer	(103)
4.6.1	Explorer 的基本功能	(103)
4.6.2	Explorer 的快捷菜单	(109)
第五章	电子图书及其检索	(112)
5.1	电子图书的特点及内容特征	(112)
5.2	电子图书目录检索	(113)
5.3	电子图书的浏览	(118)
5.3.1	收录经典文学作品的古腾堡计划	(119)
5.3.2	收录计算机信息词典的 FOLDOC	(126)

5.3.3 收录幽默笑话的 lysator	(126)
5.3.4 利用搜索引擎——“中文网探”查找电子书	(127)
第六章 电子期刊及其检索	(136)
6.1 电子期刊的特点与类型	(137)
6.1.1 电子期刊的特点	(137)
6.1.2 电子期刊的类型	(141)
6.2 电子期刊的出版商与服务	(143)
6.2.1 电子期刊的出版商	(143)
6.2.2 电子期刊的服务	(148)
6.3 电子期刊的目录检索	(152)
6.3.1 Springer 出版社的 LINK 系统	(152)
6.3.2 Blackwell's 出版商的 EJNI	(159)
6.3.3 其他一些重要的电子期刊目录检索系统	(164)
6.4 电子期刊的浏览	(168)
第七章 电子报纸及其检索	(175)
7.1 电子报纸的类型	(175)
7.2 电子报纸的目录检索	(177)
7.3 电子报纸的检索与浏览	(179)
7.3.1 人民日报	(180)
7.3.2 今日美国(USA TODAY)	(186)
第八章 Internet 网上的图书馆	(190)
8.1 图书馆的数字化革命	(190)
8.2 国家图书馆	(192)
8.2.1 北京图书馆	(192)
8.2.2 美国国会图书馆	(194)
8.2.3 英国国家图书馆——不列颠图书馆	(203)
8.2.4 澳大利亚国家图书馆	(204)
8.3 大学图书馆	(207)

8.3.1	北京大学图书馆	(207)
8.3.2	清华大学图书馆	(211)
8.3.3	武汉大学图书馆	(214)
8.4	图书情报机构	(216)
8.4.1	中国科学院文献情报中心	(216)
8.4.2	中国科技信息网	(220)
8.5	数字图书馆	(223)
8.5.1	数字图书馆的信息资源	(225)
8.5.2	数字图书馆资源和项目	(226)
第九章	Internet 信息源搜索	(245)
9.1	Internet 网络信息检索概述	(245)
9.1.1	检索工具	(245)
9.1.2	搜索索引和搜索目录的工作原理	(248)
9.1.3	Web 页面浏览技术	(249)
9.1.4	Internet 上信息质量的判断	(253)
9.2	Internet 信息源搜索方法	(255)
9.2.1	分类目录搜索	(255)
9.2.2	关键词搜索	(257)
9.2.3	搜索系统	(262)
9.3	几个重要的搜索工具	(262)
9.3.1	AltaVista	(262)
9.3.2	Excite	(266)
9.3.3	Infoseek	(268)
9.3.4	Lycos	(274)
9.3.5	Yahoo	(282)
9.3.6	搜狐(Sohoo)	(289)
9.3.7	Net Search	(291)
第十章	网络环境下电子信息服务的有关问题	(293)

10.1	信息组织方法——超文本的局限性·····	(293)
10.2	信息泛滥与信息质量控制·····	(296)
10.3	信息检索结果的处理·····	(297)
10.4	信息检索的版权保护·····	(297)
附录 1	中国图书馆网址 ·····	(302)
附录 2	世界主要图书馆网址 ·····	(311)
附录 3	中国期刊网址 ·····	(319)
附录 4	世界主要期刊网址 ·····	(322)
附录 5	中国报纸网址 ·····	(327)
附录 6	世界主要报纸网址 ·····	(332)
附录 7	中国出版社网址 ·····	(340)
附录 8	世界主要出版社网址 ·····	(345)
	参考文献 ·····	(357)
	后记 ·····	(361)

第一章 绪 论

以 Internet 为代表的网络集报刊、广播、电视的优势于一体,汇集了全球的各类信息,这些数字化信息为人类社会的发展起到了极大的推动作用,网络作为一种现代化传播工具,又为人类交流思想、传播知识提供了方便快捷的途径,成为知识经济的重要内容之一。

Internet 网先驱文顿·瑟夫于 1998 年 7 月 22 日在互联网协会的年会上说,世界 250 个国家中 240 个提供 Internet 网上网服务。约有 1 亿人使用 Internet 网,到 2000 年用户人数将达到 3~10 亿人。Internet 网以每年 100% 的速度发展,到下一个十年的中期,Internet 网的规模将超过电话系统。

据美国微波通信公司统计,1981 年美国联网的电脑不到 300 台,现在为 3000 万台以上,到 2000 年还有可能再增加一倍。据美国电脑工业年鉴公司统计,1997 年底美国 Internet 网用户总数为 5468 万。

1998 年,英国有 1000 万人网上“冲浪”,人数比 1997 年翻一番。英国市场研究组织 NOP 的一份最新调查表明,截止到 1998 年 3 月,英国 Internet 网用户已达 1030 万,占成年人口总数的 23%。调查同时显示,这些用户中近一半仅利用 Internet 网收发电子函件,真正的网上“冲浪”者人数只有 590 万左右,但这个数目却是 1997 的一倍多。NOP 的调查还发现,Internet 网目前已成为消费者购买商品前查阅参考信息的一个重要渠道。据估计,约有 200 万用户在购物之前都在网上对所要购买的商品进行过价格等方面的比较。

按美国学者的看法,一种媒体使用的人数要达到全国人口总数的 20% 以上才能被称为大众传媒。以美国为例,我们可以了解 Internet 网发展的速度。美国有 2.6 亿人口,在美国按受众达到 5000 万才算标准大众传媒的话,广播成为大众传媒用了 38 年,电视用了 13 年,有线电视用了 10 年,而 Internet 网仅仅用了 5 年。

Internet 网在中国大陆发展迅速,1996 年底互联网上用户还只有 10 万,据中国互联网络信息中心统计,到 1998 年 6 月,上网计算机为 54.2 万台,用户已达 117.5 万,网民在一年半的时间内激增了 10 倍。

Internet 网络如此迅速的发展,已大有取代传统的主要信息渠道的趋势。在信息传播领域,Internet 将引起新的变革,这些变革主要表现在:

1. 受众与媒介的关系的变革

传统的媒介如报纸、广播、电视是向受众单向地传达一种观点,灌输一种观念。受众与媒体的关系完全是单向的、灌输式的。而网络媒介则没有这种局限,它与受众的关系完全是平等的、双向的、互动的交流,且实现了文字、声音、图像的同步传输。

2. 发行与传播方式的变革

图书、报刊的消息发布是以纸为媒介,以文字为载体,靠向外界印刷发行实现传播功能。电子媒介的传播虽然可以用模拟或者数字技术达到即时传输的效果,但这两种传播方式都是单向的,观众接收是被动的。网络媒介的制作传播则完全依靠数字技术,通过调制解调器将数字信号转变成光信号,以光缆为载体传输,再将光信号转变为数字信号在另一端再现被传输的信息。这种传输方式最大的优点是传输速度快,无需节目制作的后加工。且受众可以同时发布信息的媒体进行现场交流。

3. 受众范围的变革

受图书、报刊的发行范围、传递速度的影响,其受众数量、传播

范围、影响范围是有限的。电子媒介由于近年来实现了卫星传输，有的可以远播重洋，但沟通不够。而网络媒介则不受地域、文化限制，完全实现了资源共享。受众地位也得到了提高，每个人都可以发布消息，每个人都可以成为作者、读者或评论员。

从社会层次来考虑，Internet 网络的出现，不仅仅影响了信息传播本身的变革，还对企业经营管理、经济模式乃至文化传统都将产生深远影响

罗斯扎克在《信息崇拜》中号召人类打破崇拜复归理性自信，尼葛洛庞帝则是“把比特推向大众，即使是弱小孤寂者也能发出他们的心声”，展望数字化生存为“乐观的年代”——“分散权力、全球化、追求和谐与富裕权力”是未来比特文明的“四大特质”。在未来的网络空间，文化的内涵将越来越丰富，因为在网络上为世界各地的人们提供了丰富、无限的创作思维空间，现今人们的文化创造总是局限和被动依托于地理环境、经济水平、习惯传统等，但在信息时空里，人们的平等性和依照自己兴趣作出主动选择的权力大大加强了。没有网络时，人们只能通过传统方式感受全球文化，而今天不同了，我们可以通过网络成为全球舞台上的演员，这样将有助于丰富和加强本地文化。我们将目睹地方性和全球性同时发展，地方会更加区域化，世界也具有同一性。数字化时代给我们这样的机会，这样一个发挥创造力的提升人本身智慧能力的机会。

信息技术正在潜移默化地改造我们的听觉、视觉、嗅觉、触觉、感觉，改变着人类传统的时空观念、社会运动方式、伦理道德观念和法律环境。这场席卷全球的刚刚从改变我们的行知模式开始，正以坚定的步伐深入我们生活的每一个侧面，改造着社会生活的每一个角落，更改造着人们认识世界，思考世界的观点、方法。

随着网络的不断发展，电子信息源的不断丰富，人们有理由相信，世界的发展已与网络连成一体，个人的发展和社会的进步都离不开这一切。

第二章 电子信息源

网络的快速发展,导致了电子信息源的迅速增长。信息高速公路的建成,要求“路”上有“车”,“车”上有“货”,只有这样,才能成为现代社会的一个完整的网络信息服务。

自从计算机用于信息处理以来,电子信息源从早期的书目信息,到后来的全文本信息,到今天的网络与多媒体信息,经过了 30 余年的发展,已成为现代社会不可缺少的资源财富。其处理技术从简单的字段与关键词检索,到全文检索,再到今天成为网络主流的超文本与多媒体处理。处理技术的演进与发展,一方面使电子信息源的内容从单一的书目到图文声并茂的多媒体,更富有吸引力和使用价值;另一方面,超文本的广泛利用改变传统的信息组织的线性方法,使电子信息源的组织更接近于“自然”,大大方便了用户的使用,使电子信息源的利用变得非常容易,用户不需要太多检索知识与技术。

正是由于电子信息源在内容和处理技术方面的发展,加上计算机硬件环境的变化和网络,快速、方便、有效地获取全世界的电子信息源成为了现实。

2.1 数据库

数据库是发展历史最久、影响最广的一种电子信息源,并且至今仍是一种主要的电子信息源,尽管越来越多的数据库,包括 CD-ROM 光盘数据库纷纷与 Internet 网络“联姻”,但由于其处理技术和对信息的加工深度化,在网络电子信息源快速发展的今天,它仍有发展的潜力,并将与网络电子信息源并驾齐驱。

2.1.1 美国的数据库业

美国是世界上数据库发展最早的国家。至今世界各国无论是在数据库的数量、质量、品种、类型,还是数据库生产者、数据库提供商的数量,抑或是联机数据库使用频率和产值方面,其他国家还都无法与之抗衡。据 Gale 公司统计,至 1995 年止,全世界有数据库 8525 种,其中美国产品占 69%,为世界其它国家数据库拥有总量的近 2 倍。

1. 数据库的基本概况

(1) 数据库数量

表 2.1 是 K. Y. Marcaccio 在其著作《*Computer-Readable Database: a directory and data sourcebook*》中提供的 1975 年至 1990 年美国数据库的增长情况。

表 2.1 美国数据库的增长情况

年 份	1975	1977	1979	1985	1988	1989	1990
数据库数量	177	208	259	2016	2696	3338	3982

从表 2.1 可以清楚地看到,从 1979 年开始,美国的数据库在数量上有一个飞跃。1979 年到 1985 年的短短 6 年间,数据库总数增加了 697%,从 259 个一跃超过 2016 个。而 1986 年到 1990 年几乎又翻了一番,总数将近 4000 个,占世界数据库生产数量的 60% 还多。

(2) 数据库载体和介质

数据库载体一般认为有批处理、磁带、便携式数据库、软盘、光盘、联机几种方式,而且联机数据库所占份额大大增加。1995 年,全世界社会科学、人文科学、新闻等学科领域的光盘版数据库占 71%,联机版占 18%;而发展到 1996 年,光盘版数据库所占比例下降到 56%,联机版数据库则上升至 31%。在科技、医学领域,光

盘版数据库所占比例由 1995 年的 74% 下降至 1996 年的 40%，下降了 34 个百分点；而联机版数据库则由 1995 年的 16% 上升至 1996 年的 36%，上升了 20 个百分点。由此可见，联机数据库随着全球信息高速公路的发展而逐渐占据数据库媒体的统治地位。

全文数据库在各种文字型数据库中异军突起。从 1980 年开始，全文数据库所占比例扶摇直上，从 1985 年的 28% 增加到 1995 年的 50%，其数量是书目型数据库的一倍，而书目型则从 57% 下降到 24%。另外，由于视频文字的出现和多媒体网络通信的实现，图形和图像数据库得到了进一步发展，从而改变了数据库清一色的文字信息的特点，使数据库进入了声、图、文一体化的崭新发展阶段。自 1990 年以来，图像、图形数据库的数量增长了 8 倍，比 1988 年增加了 47 倍；声音型数据库的增长势头也十分强劲，从 1988 年的 1 种增加到 1995 年的 300 种。随着超文本、多媒体和光盘驱动器的普及，多媒体数据库数量也迅速增加。

(3) 数据库主题

早年，科学、工程、技术方面的数据库曾在机读数据库中占绝对优势，到了 80 年代后期，随着个人计算机和廉价的终端开始向办公室和家庭普及，为家庭及社会服务的数据库急剧增加，数据库的主题范围也不断拓宽。时至今日，数据库几乎遍及每一个主要领域以及成百上千个细小的主题领域。

数据库主题由科技型向经济型、社会型倾斜应当说是数据库商业化、社会化的必然趋势。商贸数据库从 80 年代末期至今，在各类型数据库中一直占最大比例，目前约占 31%；科技、工程数据库占整个数据库产品的 18%；法律数据库占 12%。以科技数据库为主发展起来的 DIALOG 系统，80 年代在世界联机服务系统中独占鳌头，年联机时数达 350000 小时，占当时世界六大联机系统的 40%。但是进入 90 年代后，有迹象表明，其霸主地位已被以经济、社会信息为主的 CompuServe、America Online 和 Prodigy 所动

播。以作为面向消费者的 CompuServe 来看,它包括美食家、五号街顾客指南、好莱坞热线、音乐信息等各种服务性数据库,可谓五花八门、应有尽有。

(4) 数据库容量

美国数据库规模在发展之初即达到相当水平。1975 年美国数据库平均容量为 175 MB,而我国在 1995 年数据库容量在 10 MB 以下的尚占 33%,10—100 MB 占 42%,大于 100 MB 占 25%。80 年代,在数据库数量增加的同时,数据库的记录数以更快的速度增长。比较大的库,如世界专利数据库,容量达若干个 GB。作为一种相对更大的数据库,目前又出现了能够集合多方分散信息源中数据的数据仓库,其容量在几百个 GB 以上。

从世界范围看,在 8525 种数据库中,各类数据库规模很不平衡。有 380 种数据库的记录在 100 万条以上,其中超亿条记录的数据库有 10 种;81 种数据库的记录在 1000 万—1 亿条之间;289 种在 100 万—1000 万之间;其余数据库记录的平均值为 11.21 万条左右。

2. 数据库生产者

数据库生产者通常包括政府机构、商业/企业、非营利组织(包括学会)和综合性组织四类。由于美国存在类似欧共体的政府联合,它的数据库生产者主要由前三者构成。

政府机构曾一度是美国数据库的主要生产者。在 60 和 70 年代,大部分数据库由政府组织生产,如美国国家航空和航天局(NASA)、美国原子能委员会(现美国能源部)。作为美国信息业中举足轻重的政府机构 NTIS(美国国家技术情报服务处),至今仍是美国最大的文献型数据库生产者之一。目前 NTIS 在政府出版物交流方面已相当于一个国家级的终极市场,美国国内外许多客户已对其产品有不可替代的依赖。

1980 年以来,由于数据库商业化的冲击,数据库生产者中商