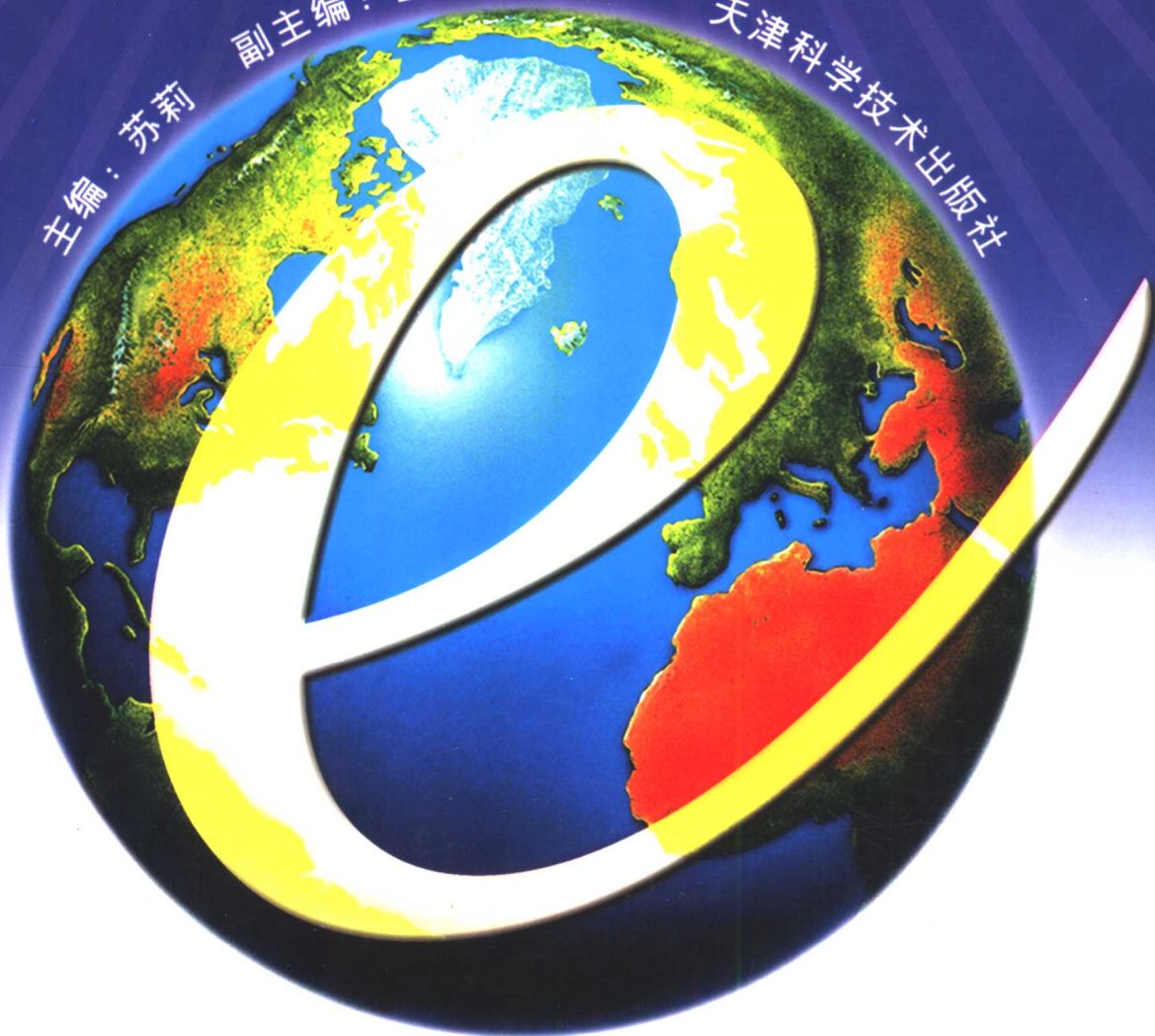


# 网络科技信息 检索

主编：苏莉 副主编：王梅 王娟萍

天津科学技术出版社



# 网络 科技信息检索

主 编 苏 莉

副主编 王 梅 王娟萍



天津科学技术出版社

## 图书在版编目(CIP)数据

网络科技信息检索/苏莉主编. —天津:天津科学技术出版社, 2003. 7

ISBN 7-5308-3454-1

I. 网… II. 苏… III. 计算机网络—情报检索  
IV. G354.4

中国版本图书馆 CIP 数据核字(2003)第 011056 号

责任编辑:王定一 吴文博  
版式设计:雒桂芬 周令丽  
责任印制:白彦生

天津科学技术出版社出版

出版人:胡振泰

天津市和平区西康路 35 号 邮编 300051 电话(022)27306314

网址: [www.tjkjcs.com.cn](http://www.tjkjcs.com.cn)

天津新华印刷一厂印刷

新华书店天津发行所发行

\*

开本 787×1092 1/16 印张 12.75 字数 295 000

2003 年 7 月第 1 版

2003 年 7 月第 1 次印刷

印数:1~1 000

定价:20.00 元

## 序 言

苏莉任主编,王梅、王娟萍任副主编的《网络科技信息检索》一书出版了。苏莉、王梅是我的学生。看到她们的著作,作为一名老教师,我感到欣慰。她们这一代人,现在是“科技文献信息检索”领域的教学和科研主力军。作为一名这个领域的老兵,我看到了今天“山花烂漫”的时刻,内心真有“丛中笑”的愉悦。

在科技文献检索领域里,这本书具有“与时俱进,开拓创新”的特点。在短短的几十年内,信息检索先后经历了手工检索、计算机脱机检索、计算机联机检索、因特网检索等几个历史发展阶段,相应的教学与研究也与日俱进。这些研究可导致知识积累,已经成为信息管理科学中重要的组成部分。苏莉等人选择这个领域的最新阶段——网络检索编著教材,确实大大更新了教学内容,同时是非常切合当今信息检索用户要求的。

网络检索这个新领域,吸收计算机研究、信息经济研究、信息服务研究、网上信息资源组织研究、数字图书馆与复合图书馆研究等方面的人才协同研究开发,因而现已出版的一些有关网络检索的著作,呈现了不同的侧重点。有的侧重于阐述计算机的应用,有的侧重于阐述其对信息经济和社会的影响,有的侧重于网络信息组织的方式方法等等。而本书的特点是从信息资源获取与信息服务手段的基点出发,向读者阐述检索的基本原理,并重点讲述了网上资源的类型和分布,信息获取的途径与工具。书中附有大量的网页,可引导读者进入文献检索的海洋。在某种意义上来说,本书可起到获取网上各类信息的实用手册的作用。本书的这种特点是作者们长期从事图书馆工作,具有丰富实践经验,了解用户需求,具备广博文献知识的体现。我相信此书一定会受到广大科技工作者的欢迎。

作为高校文献检索的教材来说,此书也有较大的贡献。20世纪70年代的此类教材,多以手工检索知识为主。80年代逐步加进了计算机检索的内容。90年代后期出现了网络检索的教材,但不够系统具体。此书在我看来,是较好的一部。此书的出版,也是对信息检索教学的一个贡献。当然,随着时间的推移,今后科技信息方面的网站、网页、搜索引擎、检索方法将有变化。我希望此书出版后,作者们要积累新的材料和知识,为今后再版做准备,以实现知识更新,保证素材的新颖性。

祝贺本书的出版!

武汉大学信息管理学院

# 前言

21 世纪,是一个网络化的新时代。随着 Internet 和 Web 技术的发展,网络环境给我们展现了一个五彩斑斓的信息世界,它不但使人们突破了时空限制;真正体会到地球已经变成了一个“地球村”;而且还改变了人们的思维和生活习惯,人们的情感也将伴随着互联网而起伏跌宕;更重要的是 Internet 引发了一场具有历史意义的数字化信息革命。

Internet 的问世,大大超出了技术范畴,也远远超出了信息产业的范围。它对经济、政治、文化、科学、教育等人类社会各个领域,都产生了重大的影响。随着近年来 Internet 的飞速发展,它也给文献信息检索带来了一场革命,使传统的手工检索跨越到网络检索的进程。

当今社会,信息资源是现代社会生产力的基本要素,是一种战略资源,信息资源的开发、利用在很大程度上决定了一个国家的经济水平、竞争实力。因此,提高全民开发、利用网络信息资源的能力已经成为当务之急。Internet 是来源最广、传播速度最快,且又分散无序的信息海洋。网络环境下海量信息的有序组织、管理与检索,无疑是目前和今后一段时间内值得我们认真关注的前沿课题。

《网络科技信息检索》正是适应了读者的这种需要,它的目的就是要为当代的大学生、研究生、教师、科技工作者和研究人员指出一条网上检索的捷径。不仅使他们懂得利用网上资源,更要学会如何利用网上资源,如何在不断奔腾、变化的信息海洋中查找本专业的信息。同时该书还是一本难得的科技信息网络检索工具。

本书的编写人员,多年从事文献信息检索的教学与计算机检索工作,始终站在情报检索的前沿,有着丰富的计算机检索和网络检索经验。该书的编撰,也是对他们多年从事检索工作的一个全面总结。

本书是由天津科技大学图书馆的苏莉构思、编撰写作大纲,经武汉大学图书馆的王梅修改并参编,天津大学的王娟萍等人合作编著,最终由苏莉审定、统稿。

在该书的编写过程中,得到了南开大学信息资源管理系主任王知津教授(博士生导师)、天津科学技术出版社王定一编审、天津大学图书馆、南开大学图书馆信息服务部同志们的关心和支持,在此一并表示衷心的感谢。

下面是参编人员以及她们所撰写的章节:

苏 莉(天津科技大学图书馆):

第一章 计算机信息检索基础

第二章 网络信息搜索引擎

第四章 国外网络综合信息资源库

4.4 Ingenta 电子期刊服务系统

第五章 国外网络信息资源库

5.1 美国《工程索引》的网络版

第八章 学位论文的网上检索

王 梅(武汉大学图书馆):

第三章 国内网络的信息资源库  
第四章 国外网络综合信息资源库

4.1 DIALOG 系统

4.2 OCLC Firstsearch 系统

4.4 Ingenta 电子期刊服务系统

王娟萍(天津大学图书馆):

第九章 网上标准文献与检索

第十章 科技报告及其网上检索

王 卉(天津科技大学图书馆):

第六章 中国专利与外国专利的网上查询

高淑萍(天津科技大学图书馆):

第七章 会议信息文献的网络检索

姚立华(天津科技大学图书馆):

第五章 国外网络专业信息资源库

5.2 美国《化学文摘》光盘数据库(CA on CD)检索

雷 霆(武汉大学图书馆):

第四章 国外网络综合信息资源库

4.3 CSA 剑桥科学文摘数据库

第五章 国外网络专业信息资源库

5.3 《英国科学文摘》(INSPEC)数据库的使用方法

白 榕(天津科技大学图书馆):

第四章 国外网络综合信息资源库

4.5 Elsevier Science 的电子期刊

苏 莉

于天津科技大学

# 目 录

|                                   |        |
|-----------------------------------|--------|
| <b>第一章 计算机信息检索基础</b> .....        | ( 1 )  |
| 1.1 概述 .....                      | ( 1 )  |
| 1.1.1 Internet 网络的发展现状 .....      | ( 1 )  |
| 1.1.2 计算机信息检索发展简史 .....           | ( 2 )  |
| 1.2 计算机信息检索原理 .....               | ( 3 )  |
| 1.2.1 基本概念 .....                  | ( 3 )  |
| 1.2.2 计算机信息检索类型 .....             | ( 4 )  |
| 1.2.3 计算机信息检索技术 .....             | ( 6 )  |
| 1.2.4 检索策略的制定与调整 .....            | ( 8 )  |
| <b>第二章 网络信息搜索引擎</b> .....         | ( 11 ) |
| 2.1 搜索引擎的含义 .....                 | ( 11 ) |
| 2.2 国际搜索引擎 .....                  | ( 12 ) |
| 2.2.1 Yahoo .....                 | ( 12 ) |
| 2.2.2 Google .....                | ( 16 ) |
| 2.2.3 Infoseek 与 Go Network ..... | ( 17 ) |
| 2.2.4 Lycos .....                 | ( 19 ) |
| 2.2.5 Excite .....                | ( 23 ) |
| 2.3 国内搜索引擎 .....                  | ( 26 ) |
| 2.3.1 Sohu .....                  | ( 26 ) |
| 2.3.2 北极星 .....                   | ( 27 ) |
| 2.3.3 网易 .....                    | ( 28 ) |
| 2.3.4 新浪 .....                    | ( 28 ) |
| 2.3.5 北大天网中英文搜索引擎 .....           | ( 30 ) |
| 2.3.6 小结 .....                    | ( 33 ) |
| <b>第三章 国内网络的信息资源库</b> .....       | ( 34 ) |
| 3.1 中国科学院文献情报中心数据库 .....          | ( 34 ) |
| 3.1.1 数据库体系 .....                 | ( 34 ) |
| 3.1.2 中西日俄文期刊联合目录数据库 .....        | ( 34 ) |
| 3.1.3 中国科学引文数据库 .....             | ( 35 ) |
| 3.1.4 中国数学文献数据库 .....             | ( 36 ) |
| 3.1.5 中国物理学文献数据库 .....            | ( 37 ) |

|                                      |               |
|--------------------------------------|---------------|
| 3.1.6 联合西文期刊篇名目次库 .....              | ( 38 )        |
| 3.1.7 中国科学计量指标数据库 .....              | ( 39 )        |
| 3.1.8 中科院科技成果数据库 .....               | ( 39 )        |
| <b>3.2 科学数据库 .....</b>               | <b>( 39 )</b> |
| 3.2.1 科学数据库简介 .....                  | ( 39 )        |
| 3.2.2 专业数据库 .....                    | ( 40 )        |
| 3.2.3 非专业数据库 .....                   | ( 41 )        |
| 3.2.4 科技导航 .....                     | ( 41 )        |
| <b>3.3 中国高等教育文献保障系统 .....</b>        | <b>( 43 )</b> |
| <b>3.4 中文科技期刊数据库 .....</b>           | <b>( 44 )</b> |
| 3.4.1 中文科技期刊数据库的简介 .....             | ( 44 )        |
| 3.4.2 中文科技期刊数据库的学科范围 .....           | ( 45 )        |
| 3.4.3 中文科技期刊数据库的检索 .....             | ( 45 )        |
| <b>3.5 万方数据资源系统 .....</b>            | <b>( 47 )</b> |
| 3.5.1 系统概述 .....                     | ( 47 )        |
| 3.5.2 数据库系统 .....                    | ( 48 )        |
| 3.5.3 科技信息子系统 .....                  | ( 49 )        |
| 3.5.4 数字化期刊子系统 .....                 | ( 52 )        |
| 3.5.5 商务信息子系统 .....                  | ( 52 )        |
| <b>3.6 中国期刊网 .....</b>               | <b>( 54 )</b> |
| 3.6.1 中国期刊网的简介 .....                 | ( 54 )        |
| 3.6.2 中国期刊网的学科内容 .....               | ( 54 )        |
| 3.6.3 中国期刊网的检索途径 .....               | ( 55 )        |
| 3.6.4 检索步骤 .....                     | ( 58 )        |
| <b>3.7 书生之家 .....</b>                | <b>( 59 )</b> |
| 3.7.1 系统简介 .....                     | ( 59 )        |
| 3.7.2 中华图书网检索 .....                  | ( 60 )        |
| 3.7.3 书生数字信息阅读器简介 .....              | ( 61 )        |
| <b>第四章 国外网络综合信息资源库 .....</b>         | <b>( 62 )</b> |
| <b>4.1 DIALOG 系统 .....</b>           | <b>( 62 )</b> |
| 4.1.1 DIALOG 的概况 .....               | ( 62 )        |
| 4.1.2 各检索平台进入步骤 .....                | ( 63 )        |
| 4.1.3 DIALOG 的检索方法 .....             | ( 64 )        |
| 4.1.4 DIALOG WEB 的检索步骤 .....         | ( 65 )        |
| <b>4.2 OCLC FirstSearch 系统 .....</b> | <b>( 70 )</b> |
| 4.2.1 OCLC FirstSearch 的概况 .....     | ( 70 )        |
| 4.2.2 OCLC FirstSearch 的数据库资源 .....  | ( 71 )        |
| 4.2.3 FirstSearch 的检索 .....          | ( 72 )        |

|            |                                 |               |
|------------|---------------------------------|---------------|
| 4.3        | CSA 剑桥科学文摘数据库 .....             | ( 76 )        |
| 4.3.1      | 数据库简介 .....                     | ( 76 )        |
| 4.3.2      | 数据库的检索方法 .....                  | ( 76 )        |
| 4.3.3      | 检索操作符 .....                     | ( 80 )        |
| 4.3.4      | 检索结果的显示和下载 .....                | ( 80 )        |
| 4.3.5      | 检索历史 .....                      | ( 81 )        |
| 4.3.6      | 常用字段介绍 .....                    | ( 82 )        |
| 4.4        | Ingenta 电子期刊服务系统 .....          | ( 83 )        |
| 4.4.1      | Ingenta 电子期刊服务系统的概况 .....       | ( 83 )        |
| 4.4.2      | Ingenta 电子期刊系统的检索方法 .....       | ( 83 )        |
| 4.5        | Elsevier Science 的电子期刊 .....    | ( 88 )        |
| 4.5.1      | 概述 .....                        | ( 88 )        |
| 4.5.2      | 检索方法 .....                      | ( 88 )        |
| <b>第五章</b> | <b>国外网络信息资源库 .....</b>          | <b>( 93 )</b> |
| 5.1        | 美国《工程索引》的网络版 .....              | ( 93 )        |
| 5.1.1      | 概况 .....                        | ( 93 )        |
| 5.1.2      | 查找方法 .....                      | ( 93 )        |
| 5.1.3      | 检索实例 .....                      | ( 98 )        |
| 5.2        | 美国《化学文摘》光盘数据库(CA on CD)检索 ..... | (100)         |
| 5.2.1      | 概述 .....                        | (100)         |
| 5.2.2      | CA on CD 的检索模式及检索途径 .....       | (101)         |
| 5.3        | 英国《科学文摘》(INSPEC)数据库的使用方法 .....  | (109)         |
| 5.3.1      | 数据库简介 .....                     | (109)         |
| 5.3.2      | 数据库使用方法 .....                   | (109)         |
| 5.3.3      | 检索结果显示与打印 .....                 | (112)         |
| <b>第六章</b> | <b>中国专利与外国专利的网上查询 .....</b>     | <b>(113)</b>  |
| 6.1        | Internet 网上专利资源概述 .....         | (113)         |
| 6.2        | 中国专利的网上查询 .....                 | (113)         |
| 6.2.1      | 全利信息网 .....                     | (113)         |
| 6.2.2      | 中华人民共和国国家知识产权局网站 .....          | (116)         |
| 6.2.3      | 中国专利信息网 .....                   | (116)         |
| 6.3        | 外国专利的网上查询 .....                 | (119)         |
| 6.3.1      | 美国专利和商标局 .....                  | (119)         |
| 6.3.2      | 欧洲专利局 .....                     | (122)         |
| 6.3.3      | WIPO 网上专利检索数据库 .....            | (123)         |
| 6.3.4      | 化学专利数据库 .....                   | (127)         |
| 6.4        | 其他可查询专利的国内外网站 .....             | (128)         |

|                                                |       |
|------------------------------------------------|-------|
| <b>第七章 会议信息文献的网络检索</b>                         | (129) |
| 7.1 概述                                         | (129) |
| 7.1.1 会议文献的特点                                  | (129) |
| 7.1.2 检索会议信息及文献的网站                             | (129) |
| 7.2 Forthcoming Conferences: 近期会议信息的网络检索       | (130) |
| 7.3 利用 OCLC 的 FirstSearch 系统中的有关数据库检索会议文献      | (131) |
| 7.4 利用 CSA 中的 Conference Paper Index 数据库检索会议文献 | (133) |
| 7.5 利用 Web of Science Proceedings(WOSP)检索会议文献  | (133) |
| 7.5.1 Full Search 的检索方法                        | (133) |
| 7.5.2 Easy Search 的检索方法                        | (135) |
| 7.5.3 常用算符                                     | (137) |
| 7.5.4 检索结果的显示、标记                               | (137) |
| 7.6 利用万方数据资源系统中的“会议文库”检索会议文献                   | (137) |
| 7.7 利用 CALIS(中国高等教育文献保障系统)检索会议文献               | (139) |
| 7.8 利用网络搜索引擎检索会议文献                             | (139) |
| <b>第八章 学位论文的网上检索</b>                           | (140) |
| 8.1 ProQuest Digital Dissertation(PQDD)        | (140) |
| 8.1.1 检索模式                                     | (140) |
| 8.1.2 检索方法                                     | (143) |
| 8.1.3 结果显示、保存与打印                               | (145) |
| 8.1.4 检索示例                                     | (145) |
| 8.2 万方数据资源系统                                   | (146) |
| 8.3 CALIS(中国高等教育文献保障系统)                        | (146) |
| 8.3.1 检索方法                                     | (146) |
| 8.3.2 检索实例                                     | (146) |
| 8.4 中国期刊网                                      | (147) |
| 8.4.1 简介                                       | (147) |
| 8.4.2 检索方法                                     | (147) |
| 8.5 北京图书馆联机公共目录馆藏查询                            | (148) |
| 8.6 中国高等学校学位论文检索信息系统(十所大学)                     | (150) |
| 8.7 国内外著名大学的网址                                 | (151) |
| 8.7.1 国内部分著名大学网址                               | (151) |
| 8.7.2 国外部分著名大学网址                               | (151) |
| <b>第九章 网上标准文献与检索</b>                           | (153) |
| 9.1 概述                                         | (153) |
| 9.2 国内标准文献及检索                                  | (153) |

|                              |              |
|------------------------------|--------------|
| 9.2.1 概述 .....               | (153)        |
| 9.2.2 国内标准数据库检索 .....        | (154)        |
| 9.3 国际标准文献及检索 .....          | (155)        |
| 9.3.1 ISO 标准 .....           | (155)        |
| 9.3.2 IEC 标准 .....           | (156)        |
| 9.3.3 PERINORM 标准数据库检索 ..... | (157)        |
| 9.3.4 联机检索数据库 .....          | (160)        |
| 9.4 发达国家标准文献网上站点 .....       | (160)        |
| 9.4.1 美国国家标准 .....           | (160)        |
| 9.4.2 英国标准 .....             | (160)        |
| 9.4.3 日本标准 .....             | (160)        |
| 9.5 专业标准(行业标准)文献网上站点 .....   | (162)        |
| 9.5.1 ASTM 标准 .....          | (162)        |
| 9.5.2 IEEE 标准 .....          | (162)        |
| 9.6 小结 .....                 | (164)        |
| <b>第十章 科技报告及其网上检索 .....</b>  | <b>(165)</b> |
| 10.1 概述 .....                | (165)        |
| 10.1.1 PB 报告 .....           | (165)        |
| 10.1.2 AD 报告 .....           | (166)        |
| 10.1.3 DOE 报告 .....          | (166)        |
| 10.1.4 NASA 报告 .....         | (166)        |
| 10.2 NTIS 数据库检索 .....        | (167)        |
| 10.2.1 国际联机检索方式 .....        | (167)        |
| 10.2.2 NTIS 数据库及检索方式 .....   | (167)        |
| 10.3 AD 报告数据库及检索 .....       | (170)        |
| 10.4 DOE 报告数据库及检索 .....      | (173)        |
| 10.5 NASA 报告数据库及检索 .....     | (177)        |
| 10.6 浏览各研究机构主页 .....         | (180)        |
| 10.7 小结 .....                | (180)        |
| <b>附录 .....</b>              | <b>(181)</b> |
| <b>参考文献 .....</b>            | <b>(189)</b> |

# 第一章 计算机信息检索基础

## 1.1 概述

计算机信息检索,是指用计算机查找文献信息的过程,它是在手工检索的基础上发展起来的。

科学技术的飞速发展,信息、情报的急剧增长,决定了手工检索已远不能适应形势的需要。人们在寻求用一种新的方法代替手工检索。计算机技术的发明,帮助人们实现了这一愿望,引起了信息检索的一场革命。计算机检索以其系统、迅速、准确等优点,日益被人们所接受。

近十年,随着信息技术、网络技术的发展,Internet 日益影响着人类的学习、工作和生活。它不仅给人们带来方便、快捷的交流方式,而且加速了信息的传递,使“资源共享”的梦想成真。因特网已成为真正知识与信息的海洋。人们可以从网上方便、快捷地获取信息。不仅可以获得新闻、通讯等信息,也可以获取专业、科技信息。如何掌握计算机检索知识,查找网络的信息资源,已成为每个科技工作人员、专家、学者、信息情报人员亟待解决的问题。

### 1.1.1 Internet 网络的发展现状

因特网从 1968 年美国国防部高级研究计划署开始研究,到 1992 年实际运行,只用了二十多年时间。尤其 20 世纪 90 年代后,因特网的发展迅猛异常。目前,已有 193 个国家和地区在因特网上注册了域名,1 600 万台计算机连入互联网,网络用户已达 1.6 亿。而且,连入 Internet 计算机的数量每年翻一番,互联网的发展从根本上改变世界的面貌,它也是人类进入信息时代的重要标志。

因特网是一个理想的信息交流媒体,人们在網上建立电子论坛、新闻论坛、电子公告栏,网络用户之间可以沟通信息,讨论问题。

因特网的发展,也促进了电子出版业的兴起。它的代表是 E-print Archives (电子印刷文档)。E-print 于 1991 年 8 月由美国 Los Alamos 国家实验室提出,开始仅为理论物理、高能物理领域的 160 个用户服务,后发展到经济、计算与语言等其他领域。一些科学家和图书馆员跳过传统的出版商,直接把论文投入 E-Print 系统。在 E-Print 系统中,论文的投寄、储存和利用服务成为一体。E-Print 的增长是爆炸性的。现有 25 000 个活跃用户,每天有 45 000 笔信息流量。这种形式奇迹般地加快了研究结果的交流。

一些出版商和联机检索系统的经营者,认识到因特网的快速、廉价的优越性带来的巨大吸引力,开始把电子出版物(电子期刊、电子图书、电子报纸)和一些著名检索刊物数据库直接输

人到因特网,如世界著名的期刊《Times》《Forbes》《Science》《Nature》《华尔街日报》《纽约时报》,我国的《人民日报》《经济日报》以及电子辞典、百科全书。世界著名检索刊物《美国工程索引》《科学引文索引》《英国科学文摘》,也在因特网上推出了自己的网络版。世界上将近 1 万多个联机检索系统的数据库也与因特网联网。大的联机检索系统(DIALOG, STN, OCLC)也纷纷在网上开辟了网页,开始了 Web 页面检索。

网上数字图书馆的产生,也是因特网迅速发展的重要标志。数字图书馆是电子出版物和数字化信息的集合体。它把数字化的馆藏(例如光盘、机读磁盘)、商用的光盘数据库和联机数据库、因特网上的信息资源搜集起来,供用户查找。它不同于传统的图书馆,是一种相关数字化信息的集合;亦不同于数据库,是数百数千计数据库的整合,是“虚拟图书馆”和“无墙图书馆”。例如,WWW. Virtual Library: Biosciences,是由美国哈福大学生物方面多个机构联合建立的,它把生物科学及其 31 个相关领域的新闻、生物科学家通讯录、与生物科学相关的 Gopher 站点目录和索引、生物学期刊目录、生物学软件、FTP 文件库及与生物学研究、教育有关的资源和工具组织到一起,实际上它是一个生物科学资源目录。

数字图书馆的概念是由美国国家科学基金会等三个机构联合发起的数字图书馆工程中首次提出的,后美国计算机界、情报图书馆界纷纷投入对数字图书馆的研究,作为信息高速公路的重要资源来建设。它不仅在美国,而且在世界各国均得到迅速发展。

### 1.1.2 计算机信息检索发展简史

从第一个科技文献检索系统算起,计算机信息检索主要分为四个发展阶段:

#### 1. 脱机检索阶段

20 世纪 50 年代初期到 20 世纪 60 年代,由于受当时计算机技术条件的限制,计算机的主要运算部件是电子管,存储介质是磁带、磁鼓。人们把信息文献存储在这些磁带、磁鼓上,组成了文档和数据库。检索系统由一台计算机和相关文档或数据库组成。专职操作员根据用户的需要,在文档和数据库中自动寻找所需文献。这一阶段,被称为脱机检索阶段。虽然脱机信息检索十分原始,但当初运用的布尔逻辑等原理,至今仍是最先进的计算机检索系统的基本原理。

#### 2. 联机检索阶段

20 世纪 60 年代中期,以半导体为主要器件的计算机出现了。1962 年,美国麻省理工学院进行了世界最早的联机情报检索实验。20 世纪 70 年代后期,随着卫星通讯技术、微机和数据库的迅速发展,美国和一些西欧国家实现了国际联机检索。用户利用终端设备,通过国际卫星通讯网,与世界各国的大型计算机检索系统的中央计算机进行人机对话,检索世界各国数据库的大量信息资料。用户还可根据检索输出情况,随时修正检索策略。远程、实时、检索多种数据库是联机检索的主要优点。但国际联机检索代价昂贵,妨碍了国际联机的普遍使用。

#### 3. 光盘检索阶段

20 世纪 80 年代中后期,出现了光盘数据库系统。光盘低廉的价格,和不受时间限制地在有光盘驱动器的计算机上实现信息检索,使它一问世就得到快速普及。20 世纪 80 年代末出现了光盘网络。光盘塔和局域网技术的应用,使多个用户检索同一大型数据库,检索效率得到很大提高。在 DOS 平台上实现多用户资源共享、存储介质的高密度化、数据存储高速化、检索费用低以及光盘数据库与联机系统共存,是这一阶段检索技术的主要特点。

#### 4. 网络检索阶段

20 世纪 60 年代末和 70 年代初,美国国防部开始了 ARPANET 的研究,这就是 Internet 研究的开始。90 年代初,因特网在全世界得到异常迅猛的发展。因特网的核心技术是 TCP/IP 协议,它解决了世界上各种不同计算机网络的相互通信问题。伴随着计算机网络的是 Windows 操作系统、超文本检索系统的出现以及微机、网络技术的普及。在这种形势下,各种情报检索系统的运营商一方面逐步升级检索软件,一方面纷纷将数据库放到因特网,以吸引更多的用户。这些都促使检索费用不断下降,检索系统功能不断增强。因此,我们可以说,操作平台 Windows 化、数据库超文本化以及通讯手段因特网化,是当前检索技术发展的主要特征。

## 1.2 计算机信息检索原理

计算机检索,从开始的脱机检索,发展到网络检索,从简单的批处理到多媒体技术、全文检索、超文本检索技术的全面应用,检索技术可谓发展神速。但万变不离其宗,无论是原始的脱机检索,还是图像声色并存的网络检索,都是以计算机检索原理为基础的。欲掌握网络检索知识,必须从掌握计算机检索基础原理开始。

### 1.2.1 基本概念

#### 1. 信息、知识、情报

人类社会已步入了信息时代。信息在自然界、人类社会和人类思维活动中普遍存在。什么是信息?它具有何种属性?

信息是表示物质及其运动的某种特征的信号系统。其中,直接地表征物质及其运动的某种特征的信号,称为“自然信息”;人类用语言、文字等符号间接地表述物质及其运动的某种特征的高级信息,称为“人工信息”或“社会信息”。

图书信息领域的专家把信息视为可以以各种形式进行传播、记录、出版及发行的观念、事实及论著。这种见解,是比较切合信息时代的需要的。

信息借助于某种载体存在,借助某种媒体传播,它具有实效性、传递性、可扩散性、可替代性和共享性等特点。

知识是人类对客观事物规律的认识。它是人们在社会实践中积累起来的经验。人们对事物由表及里,由现象到本质,由感性到理性的认识深化,便形成了知识。知识是信息内容的组成部分。

情报是在一定时间内,为了解决一个特定的问题所需要获取的特定的信息和知识。它隶属于信息的范畴之内。

情报具有三种属性。①及时性:又称新颖性。情报必须在一定时间内发挥作用,超过这时间,情报就失去价值,我们亦把它称为情报寿命。②传递性:是指信息只有通过传递才能为人们解决特定的问题,信息不传递、不更新就不能形成情报。情报是动态的,它只有传递给用户,并为用户所利用时才成为情报。③实用性:是指情报可以为人们解决特定的问题。

信息、知识、情报已成为当今社会巨大的、无形的财富,成为一种与能源、材料并存的重要战略物资。

#### 2. 信息检索定义

信息检索有两方面的含义。一方面是将信息按一定的方式、方法有序地组织起来,另一方面是根据用户的需要,将特定相关信息准确地查找出来。因此,它的全称应是信息的存储与检

索。一般来讲,信息检索是指它的后半部分,即在信息海洋中查找所需信息的过程。当前,更普遍的术语是信息存取(Information access)。“Access”一词源于计算机领域,指计算机访问文档及数据库的方式,引入信息检索领域,则拓宽了检索的内涵及其应用。信息存取将所有信息的组织、检索活动及其先进的技术手段融合在一起。

信息检索按存储媒体和技术手段可划分为手工信息检索和计算机信息检索。

计算机信息检索,是指用计算机查找所需信息的过程。它的实质是“匹配”。一方面是用户的信息需求,一方面是有顺序的信息集合。计算机检索,即由检索者把用户提问变成机器可识别的检索策略输入计算机中,由计算机自动对数据库中各文档的记录进行扫描、匹配,把完全匹配的文献记录检索出来的自动化过程。它是在手工检索的基础上发展、演化出来的。

### 3. 记录、文件/文档、数据库

(1)记录(Record) 在计算机信息检索中,每一篇加工过的存储于计算机的文献称为一条记录。在计算机数据库管理中,把磁盘中一次读写数据的基本单位也称为记录。

(2)文件/文档(File) 指一组相关记录的集合。例如,在计算机检索中经情报专家处理的文献被输入,并按照一定的格式顺序存放于计算机中,形成顺序组织的数据文件,有人称为“顺序文件”,有人称为“顺排文档”。从顺序文件中抽取相关的数据项目(如主题、作者),并把这些数据项目排序,排序后的数据文件称为“索引文件”或“倒排文档”。这里文件、文档实质上是同一种概念,人们的习惯称呼不同罢了。在人们的理解中,还有一个不同点即:文件是单一的,文档是若干文件的组成。

(3)数据库 按照国际标准组织标准 ISO/DIS5127 规定,数据库是指至少由一种文档(file)组成,能满足特定目的和特定功能数据处理系统需要的数据组合。

数据库的本质是可以提供共享的有一定组织方式的相关数据。首先,数据库储存的是一系列相互关联的数据,例如,某一学科的文献信息的集合,某一范围的物质性质研究数据的集合等。这些数据是经过调查、实验、统计、整理和归纳后得到的,而且每一类数据都有它们的共性和相互关联性。另外,为了便于管理和处理这些数据,当这些数据存入数据库时,就已具有一定的数据结构和文件组织方式。这样存入的数据就可以为多个用户反复使用,达到资源共享的目的。

(4)字段与子字段 字段(Field):在一篇文献中的每一特定数据项称为一个字段,一条记录由若干字段组成。典型的文献检索系统常有这样几个基本字段:题名、作者、单位、出版物名、文摘、分类、主题词。有些加工复杂的系统字段也多。一种字段被称为一种检索途径。

(5)子字段(Subfield) 一个字段往往由若干子字段组成,例如一篇文献有多少个作者,就意味着有多少个子字段。一篇文献往往要标引若干个主题词,也就有若干个主题子字段。

(6)主题词、标识词、叙词、关键词。

### 1.2.2 计算机信息检索类型

#### 1. 联机检索

联机检索是指用户利用计算机终端设备,通过通信线路,从信息中心的主机数据库中检索出所需要的信息的过程。它允许用户以人机对话、联机会话这样交互的方式直接访问系统及数据库。用户的提问一旦传到主机并被接收后,机器便立即执行检索运算,很快将运算结果传送到用户终端,用户可反复修改检索式,最后获得较满意的检索结果。

联机检索能远程登录到国内外检索系统,大型国际联机检索系统的数据库大概共有一万

个。大型检索系统不仅数据库多,而且数据库的文献报道量大,数据更新及时,系统检索点多,组合方式多样,输出形式、输出方式多样。用户容易得到最新、最准确和最完全的检索效果。

联机通信方式有拨号入网、租用专线等。许多联机检索系统可以通过互联网访问。

联机检索传统上使用命令式检索,通过一系列能够识别的命令(指令),完成对检索式的运算,实施检索。命令方式能够正确表示检索词之间的概念,表达检索词之间的关系。检索快速、简捷,检索精确度高,但由于检索价格昂贵,需要由熟悉检索命令、检索技能高的专职人员执行,一般科技人员不能胜任。

## 2. 光盘检索

光盘应用计算机技术、激光技术、多媒体技术存取数字信息。由于使用激光高密度存储技术,致使一直径长 120mm 的普通光盘,信息存储量高达 650M。因光盘具有轻便、易携带、耐用等特点,从上世纪 80 年代中期后便很快占领市场,广泛用作信息载体、检索工具。

检索用光盘均为“只读式光盘”(Compact Disc-Read Only Memory,简称 CD-ROM),它的数据只能被读取,一般电子出版物的载体多为 CD-ROM。

光盘数据库用于检索的有两类:单机检索系统和联机检索系统。

(1)单机光盘检索系统 该系统由微机、光驱、光盘数据库等硬件设施,以及检索程序、驱动程序、操作程序等软件组成,自成一体,提供单机、单用户使用,系统结构简单,但数据量少,利用率低。

(2)联机光盘检索系统 该系统只提供局域网上的检索,如图书馆网、校园网等。它可以连接到许多用户终端,网上用户可以分时共享光盘数据库信息。光盘有服务器管理运行,它们的光驱是多盘的,有塔式光驱和自动光盘换盘机,即光盘塔和光盘库。

光盘检索通常采用菜单方式,根据菜单提示、指引,通过选择、确定和键入填写,以及一些功能键的使用,一步步地执行检索,修改检索提问,直至完成全过程。光盘检索操作简单,使用方便,其缺点是检索步骤多,操作繁琐,检索精度一般也不如命令式检索。

## 3. 网络检索

它具有全球式的分布结构、开方式的信息环境及跨国界的信息流。互联网上的信息源浩如烟海,诸如联机数据库、网络数据库、电子出版物、数字图书馆、数以万计的网站等等。先进的搜索技术,使得用户不用知道网址,就可查到所需信息。这就极大扩展了检索空间和信息量。而传统的联机检索和光盘检索,只局限在一台或几台计算机的特定数据库的检索上。但另一方面,互联网信息量过大,信息庞杂,缺乏有效的组织管理,因此,很难用一般意义上的查准、查全这些概念来衡量。

网络检索主要采用超文本检索和混合检索。

(1)超文本检索 超文本是用文件里的词、词组、图像将文件连接起来,通过这些链可以从一个文件转到另一个文件。超文本的内容排列是非线性的,它按知识(信息)单元及其关系建立起知识结构网络。具有图形的信息又称超媒体,超文本/媒体的检索是通过超文本(媒体)链接来实现的。其形式有的在网页的文字处有下划线,或以图标方式标志,用户点击这些标志便能进入到与此信息相关的下一页,在该页面上通过超文本链接进入再一个界面,用户通过这些页面获取自己所需要的信息。

(2)混合型检索 混合型检索多见于互联网 WWW 版本的文献检索。WWW(World Wide Web)即 Web,它是用 WWW 浏览器在 Windows 界面下交互作业,能给用户揭示到一篇篇文献的

信息,具有很强的直观性。

Web 文献数据库检索系统是在大量采用超文本的基础上又将命令检索、菜单检索方式融合其中,交互使用,集所有检索机制为一体,许多大型国际联机检索系统(Dialog、STN、OCLC)在互联网上设有自己的站点,提供用户方便的检索服务。

### 1.2.3 计算机信息检索技术

#### 1. 字段检索

字段检索是限制检索词在数据库内查找范围的一种计算机检索技术。文献数据库由若干文献记录组成,每条文献记录含有若干字段:主题词、叙词、关键词、文献题目、作者、作者所在工作单位、期刊名称、文献类型、语种、出版机构、出版年、分类代码、专利号、科技报告号、登录号等。确定了检索字段,就等于确定了检索途径,这有利于缩小检索范围,提高查准率。

各数据库基本检索字段的内容和数量均不相同,字段标示符号亦不相同,所以在使用前,必须参考各数据库的使用说明。

#### 2. 布尔逻辑检索

当用户检索某个课题时,一般不只用一个检索词而是用多个检索词来表达他的主题内容,这就需把检索词进行合理的逻辑组配。布尔逻辑算符就是用来组配这些检索词和代码的。布尔逻辑检索是计算机检索的基础。

常用的布尔逻辑算符有:逻辑与(AND);逻辑或(OR);逻辑非(NOT)。

(1)逻辑与(AND) 逻辑与用算符 AND 表示。A and B,表示既含有 A 又含有 B。例如,solar and energy,表示同时含有 solar 与 energy 的文献。检索时同时含有 solar 与 energy 的文献才算命中。这种逻辑关系如图 1.1 所示。使用逻辑与,增强了检索的专指性,缩小了检索的范围。

(2)逻辑或(OR) 逻辑或用算符 OR 表示。A or B,表示或含有 A,或含有 B,或同时含有 A 与 B。例如:solar or sun,表示一篇文献只要含有一个词就算命中。这种逻辑关系如图 1.2 所示。逻辑或扩大了检索范围。

(3)逻辑非(NOT) 逻辑非用算符 NOT 表示。A not B,表示含有 A 而不含有 B。例如:energy not solar,表示检出文献中,必须含有 energy 不能有 solar(图 1.3)。逻辑非与逻辑与一样,缩小了检索的范围,增强了检索的准确性。

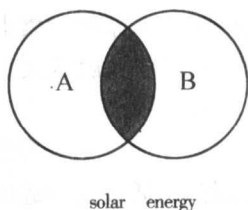


图 1.1 “逻辑与”示意图

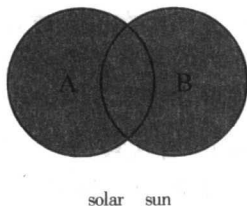


图 1.2 “逻辑或”示意图

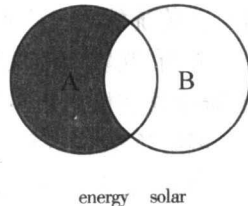


图 1.3 “逻辑非”示意图

对于一个布尔逻辑检索式,检索系统处理的顺序是从左到右执行。一个检索式同时有 and, or 和 not 时,处理顺序是:先执行逻辑非,再执行逻辑与,最后执行逻辑或。检索式里有括号时,先运算括号内的程序,如有多个括号,最先运算最内层括号的程序。

#### 3. 截词检索

截词检索的目的是把与提问词中相同词干的文献检索出来。它广泛应用于西文文献检索