

主 编 安邦建
副主编 钟哲辉 王志斌
李冕斌 黄晓英

计算机 网络 信息 检索

北京师范大学出版社

G252.7
200206
A

计算机网络信息检索

主 编：安邦建

副主编：钟哲辉 王志斌

李冕斌 黄晓英

按照课第17项的“检索提问单”的格式和步骤完成你
选择的课题在《中国学术期刊全文数据库》的检索。并根
据检索结果说明那些信息检索的基本方法和对查询
列 检索查全率和查准率

要打印出检索结果的题录并选择相关性最大的一篇文章
的全文。

经济全球化和中国法制建设



* A 0 8 8 0 9 1 9 *

北京师范大学出版社

北京师范大学出版社出版发行
(100875 北京新街口外大街 19 号)

出版人:常汝吉

海南师范学院印刷厂 全国新华书店经销

开本:850×1168 1/32 印张:11.25 字数:290 千

2002 年 6 月第 1 版 2002 年 6 月第 1 次印刷

印数:1—3000 册

定价:20.00 元

内 容 提 要

本书在坚持理论联系实践的基础上,详细阐述计算机网络信息检索的基本概念、基本原理、基本技术和方法,按照计算机网络信息的不同载体形式,系统介绍各种计算机网络信息资源的基本特征、信息价值、各种中外文检索平台以及对它们进行有效检索与利用的方法。

本书具有创新性、系统性、知识性及实用性等特点。既可作为大中专院校的通用教材,也可作为参考工具书、教学参考书。适用于信息技术开发与利用、信息管理与信息系统及相关专业的学生、教师用书,广泛适用于科技工作者、以及所有需要从网络中获取信息的人员。

序 言

海南大学安邦建等先生编著的《计算机网络信息检索》一书出版了。面对这本专论网络检索的著作,我是感到欣慰的,并乐于为之写序。

有关信息检索方面的教材和论著,自上世纪七十年代初出现以来,已经历了几个发展阶段。大致说来,七十年代出版的著作,基本上是阐述手工检索的理论和实践,其间只涉及穿孔卡片检索和缩微胶卷检索等简单的机械检索和光电检索的知识。八十年代,随着计算机信息检索服务的发展,特别是国外联机检索信息服务系统的终端伸延到我国之后,信息检索方面的论著在内容上有了实质的发展,即除了包含手工检索之外,又论述了计算机检索——包括联机批处理检索和联机检索的理论与方法。而且随着时间的推移,计算机检索知识所占的比重日趋增强。开始是以手工检索为主,计算机检索为辅;其后,两者平起平坐、平分秋色。再后来,则计算机检索超过了手工检索的份量。这一变化,改变了高校开设的“文献检索”课程的面貌,提高了它在高校课程体系中的地位,并赢得了社会对这一学科分支的更加尊重。九十年代,随着因特网的发展,信息检索方面的论著又增添了因特网检索的内容,以至使手工检索、联机检索和因特网检索三者的知识融为一体。这种与日俱进的内容范围和知识体系的扩充与发展,使信息检索这一学科领域充满了时代信息和充沛的生命力。

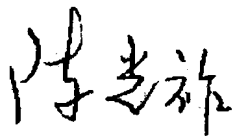
今天,时至廿一世纪,安邦建等先生的《计算机网络信息检索》一书,则以网络检索(特别是因特网检索)知识的论述为自己主要的任务,当然也涉及了光盘检索和 Dialog 等联机检索。这种内容

重点的选择,反映了这一学科发展的新面貌,也适应了读者的当前的重点需求。因特网的发展影响了当代生活的各个方面,其中包括了人们获取信息的行为和方式。因此,本书的出版,是体现了这一时代潮流的。我想它将获得读者的欢迎。

本书的另一特点是简洁性。全书文字不多,但重点突出,文字较为简炼。而且本书包括了许多图表和实例,便于读者举一反三、按图索骥。从而达到授人以渔的目的。

当然,本书中的某些部分显得较薄弱,相信在本书再版时将变得更充实。

最后,我想说的是,本书作者们多系青年人,虽然他们的著名度不高,但勇于探索学科前沿的精神是可嘉的。这一点,也是我作为一个信息检索领域的老兵感到高兴的。我期待着他们以及其他青年学人迅速成长。

A handwritten signature in black ink, consisting of three characters: '陈' (Chen), '志' (Zhi), and '群' (Qun), written in a cursive style.

2002年2月24日于武汉大学信息管理学院

前 言

1969 年阿帕网(ARPANET)问世以来,计算机网络技术得到了突飞猛进的发展,将人们带进了一个全新的网络时代。随着 Internet、Intranet 技术和知识经济的发展,网络信息资源呈爆炸性的递增态势,为此,人们就如何有效地检索计算机网络信息进行了积极的探索。

为了满足人们对计算机网络信息检索相关知识和技术的迫切需要,各种相关的书籍如雨后春笋般地涌现出来,使读者应接不暇。但是,在已出版的书籍中,往往只是针对电子资源的检索与利用,缺乏对计算机网络信息资源的研究、检索和利用,尤其缺少对计算机网络信息检索的检索原理进行有益的探索和研究。基于此,我们组织了一批长期从事计算机网络信息检索方面的专业人员,根据他们多年从事计算机网络信息检索教学、研究和开发的实践,通过对计算机网络信息前沿科技的跟踪研究,完成了本书的编著工作。

本书在坚持理论与实践相结合的基础上,主要阐述计算机网络信息检索的基本知识、基本概念、基本原理、技术和方法,按照计算机网络信息不同载体形式,系统介绍各种计算机网络信息资源的基本特征、信息价值以及对它们进行有效检索的方法。第一章阐述计算机网络信息检索系统及其检索原理;第二章、第三章分别介绍 Internet 信息检索和中国互联网络信息检索的基本概况、相关网络结构、信息资源的检索与利用;第四章介绍 Dialog 联机信息检索;第五章介绍网络光盘数据库信息检索;第六章介绍网络书目数据库检索;第七章介绍多媒体信息检索。

本书第一章由安邦建执笔;第二章、第三章和第五章由钟哲辉执笔;第四章由王志斌执笔;第六章由李冕斌执笔;第七章由黄晓英执笔。全书由安邦建修改定稿。

由于作者水平有限,成稿时间较仓促,错误之处在所难免,敬请广大读者批评指正,提出宝贵意见,以便我们进一步完善。

作 者

2001年9月10日

目 录

序	(1)
前言	(1)
第一章 计算机网络信息检索系统概述	(1)
第一节 计算机网络信息检索基本原理	(1)
一、计算机网络信息检索及其发展	(1)
二、计算机网络信息检索基本原理	(3)
三、计算机网络信息检索的主要特点	(4)
四、制定检索策略的步骤和方法	(6)
第二节 计算机网络信息检索系统构成及搜索引擎 ..	(10)
一、计算机网络信息检索系统结构	(10)
二、计算机网络信息检索系统中的人工智能技术	(13)
三、搜索引擎概述	(20)
四、有待研究的关键技术	(24)
第三节 计算机网络信息检索工具介绍	(25)
一、交互式检索工具	(26)
二、字典型检索工具	(29)
三、索引型检索工具	(31)
第二章 Internet 信息检索	(35)
第一节 Internet 概述	(35)
一、TCP/IP 协议	(36)

二、DN 和 IP 地址	(37)
三、URL	(39)
四、访问 Internet 的方式	(40)
五、Internet 的主要功能	(41)
第二节 WWW 信息检索	(42)
一、WWW 信息检索概述	(42)
二、WWW 信息检索与应用	(45)
第三节 Gopher 信息检索	(55)
一、Gopher 概述	(55)
二、Gopher 信息检索与应用	(60)
第四节 WAIS 信息检索	(70)
一、WAIS 概述	(70)
二、WAIS 信息检索与应用	(73)
第五节 Archie 信息检索	(83)
一、Archie 概述	(83)
二、Archie 信息检索与应用	(86)
第六节 FTP 信息检索	(98)
一、FTP 概述	(98)
二、FTP 信息检索与应用	(102)
第三章 中国互联网络信息检索	(111)
第一节 中国互联网络概述	(111)
一、网络建设	(112)
二、资源和组织规范建设	(115)
第二节 CHINANET 信息检索	(119)
一、Chinanet 概述	(119)

二、Chinanet 信息资源	(121)
三、Chinanet 信息检索与应用	(124)
四、Chinanet 信息导航中心	(126)
第三节 CERNET 信息检索	(130)
一、CERNET 建设原则和技术特点	(131)
二、CERNET 网络结构	(133)
三、国家网络中心和地区网络中心	(135)
四、入网方式及其应用	(137)
五、CERNET 信息检索与应用	(139)
第四节 CSTNET 信息检索	(144)
一、CSTNet 概述	(144)
二、CSTNet 网络结构	(145)
三、CSTNet 网络构成	(149)
四、CSTNet 信息检索与应用	(151)
第五节 CHINAGBN 信息检索	(155)
一、ChinaGBN 概述	(155)
二、ChinaGBN 网络资源	(159)
三、ChinaGBN 信息检索与应用	(162)
第四章 Dialog 联机信息检索	(167)
第一节 Dialog 联机检索概述	(169)
一、Dialog 基本检索方法	(169)
二、Dialog 系统检索指令	(189)
三、Dialog 系统检索实例	(201)
第二节 Dialog 系统常用数据库简介	(203)
一、INSPEC	(203)

二、Ei COMPENDEX* Plus	(205)
三、CA SEARCH	(207)
四、ISMEC	(208)
五、METADEX	(208)
六、DISSERTATION ABSTRACTS ONTINE	(209)
七、ENVIROLINE	(209)
八、POLLUTION ABSTRACTS	(210)
九、TRIS	(210)
十、PASCAL	(210)
十一、Peterson's Gradline	(211)
十二、ENGINEERED MATERIALS ABSTRACTS	(211)
第五章 光盘数据库信息检索	(213)
第一节 KR OnDisc 检索系统与 Ei 光盘数据库信息	
检索	(213)
一、KR OnDisc 菜单式检索与应用	(213)
二、KR OnDisc 命令式检索与应用	(217)
第二节 ProQuest 检索系统与 SA、INSPEC 光盘数据	
库信息检索	(221)
一、ProQuest 的基本检索	(222)
二、ProQuest 的字段检索方法	(225)
三、ProQuest 的功能键、逻辑运算符及截词符	(225)
四、SA 检索及其应用	(227)
五、INSPEC 数据库检索及其应用	(229)
第三节 《Chemical Abstracts》光盘数据库信息检索 ...	(236)
一、CA on CD 概述	(236)

二、CA on CD 检索与应用	(237)
第四节 OCLC 光盘数据库信息检索	(240)
一、OCLC 概述	(240)
二、OCLC 光盘数据库信息检索及应用	(243)
第五节 《中国学术期刊》全文数据库信息检索	(244)
一、《中国学术期刊》概述	(244)
二、《中国学术期刊》全文数据库信息 检索及应用	(245)
第六节 《中文社科报刊篇名数据库》信息检索	(253)
一、《中文社科报刊篇名数据库》概述	(253)
二、《中文社科报刊篇名数据库》信息检索及应用	(254)
第七节 《中文科技期刊篇名数据库》信息检索	(261)
一、《中文科技期刊篇名数据库》概述	(261)
二、《中文科技期刊篇名数据库》信息检索及应用	(261)
第六章 计算机书目信息检索	(271)
第一节 书目、书目信息检索与检索工具	(271)
一、书目的作用与类型	(271)
二、书目检索途径	(273)
第二节 图书馆自动化集成系统书目信息检索	(288)
一、图书馆集成系统功能简介	(288)
二、图书馆集成系统书目信息检索与应用	(289)
第三节 计算机网络书目信息检索	(298)
一、CERNET 书目信息检索	(299)

二、中国期刊全文数据库检索	(304)
第七章 多媒体信息检索	(309)
第一节 多媒体概述	(309)
一、什么是多媒体	(309)
二、多媒体信息的特性	(311)
三、多媒体信息系统	(312)
四、多媒体数据	(313)
五、多媒体数据的特性	(315)
六、多媒体数据库管理系统	(316)
七、多媒体信息检索与查询	(318)
第二节 基于内容的检索	(319)
一、基于内容的检索及其特点	(319)
二、基于内容检索的关键技术	(320)
三、基于内容检索的体系结构	(323)
四、基于内容检索分类	(326)
五、基于内容的检索过程	(326)
六、基于内容的信息检索方法	(327)
第三节 基于内容的多媒体信息检索与 MPEG7	(341)
一、MPEG7 与基于内容的多媒体信息检索的联系	(341)
二、MPEG7 的主要内容	(342)
三、MPEG7 的应用范围	(343)

第一章 计算机网络信息检索系统概述

计算机网络是实现计算机硬件、软件和电子信息等资源共享的基础设施。通过计算机网络,用户不但可以使用本地计算机资源和电子信息源,还可以通过网络访问联网的远程计算机资源和电子信息源。目前,计算机网络以 Internet 和 Intranet 为标志,以 WWW(World Wide Web)超文本工具为依托,推动信息检索系统和信息检索飞速发展。

第一节 计算机网络信息检索基本原理

一、计算机网络信息检索及其发展

Internet 的前身是 ARPANET 和 NSFNET。1983 年 1 月 TCP/IP 协议成为正式的美国国防部计算机网络 ARPANET 的网络协议标准后,大量的网络、主机和用户都连入了 ARPANET,使得美国国防部高级研究计划局(ARPA, Advanced Research Project Agency)资助的 ARPANET 迅速发展。当美国国家科学基金会(NSF, National Science Foundation)资助的 NSFNET 取代了 ARPANET 后,这种发展呈指数式增长,成为今天的互联网(Internet)。进入 80 年代中期,人们开始认识这种大型互联网络的重要作用。网络信息基本服务主要包括:电子邮件(E-mail)服务、文件传输(FTP)服务和远程登录(Telnet)服务等。

众所周知,Internet 是一个覆盖全球的“网络的网络”,其现有

规模已经大得令人瞠目结舌。根据 1999 年 3 月的统计数字,世界上已经有大约 5,000 万个大小不同的计算机网络与 Internet 相连,上网的计算机总数已达 12,000 多万台之多,而且与 Internet 联网的计算机台数仍在以惊人的速度不断增长(根据中国互联网信息中心的公报 1999 年第 2 期)。实际上,每一台计算机都是一个信息源,存贮在其上的文件(或数据库)若平均以 30 个计,则文件(或数据库)的总数将高达 36 个亿以上。当然,一台计算机上的信息资源绝不仅如此。因此,Internet 的信息资源是名符其实的“信息海洋”。用户在如此浩瀚的“大海”中寻找自己需要的信息,无异于“大海捞针”。

于是一类被人们称为网络信息检索工具(Networked Information Access Aid)的新型检索工具应运而生。网络信息检索的研究和网络信息检索工具的开发使原来的“情报检索”这一学科分支产生了新的分支学科“网络信息检索”(NIR, Networked Information Retrieval)。

信息检索和信息检索工具对理论界来说,并不是什么陌生的词汇。70 年代,当联机信息检索系统开始投入使用以来,人们已经开发了许多种类的情报(信息)检索软件,形成了软件开发研究和情报科学研究的一个相对独立的分支:信息检索。尽管信息检索软件的功能与日俱增,技术水平不断提高,但是这些软件的运行环境基本上停留在单机或局部网络。80 年代虽然出现了分布式的信息检索系统(如分布在美、德、日三国的 STN 系统),但其主机数量十分有限,只是分布在几个点上。

随着 Internet 的不断扩张,传统的信息检索手段变得越来越困难,要求信息检索工具的运行环境已经不再是单机或连接为数不多几个点的局部网络(LAN, Local Area Network),而是几乎覆盖全球的广域网(WAN, Wide Area Network)。因此,计算机网络信息检索已经不是原来意义上的情报(信息)检索,而原有的情报(信息)

检索工具已不能适应新的网络环境的需要。

Internet 网上信息资源的迅猛增长使主管 Internet 技术问题的 Internet 工程特别工作小组 (IETF, Internet Engineering Task Force) 对网络信息检索这一重要问题给予了充分重视。IETF 与另外两个网络机构即欧洲科研网络协会和网络信息联盟联合建成“网络信息检索工作小组”(NIR Working Group)。网络信息检索工作小组的主要任务是:

1. 收集和发布有关网络信息检索工具及其研究团体的信息;
2. 鼓励合作开发网络信息检索工具;
3. 协调网络信息检索工具开发单位、有关联网机构和网络信息检索工具的生产销售厂商的行动。

1994 年 8 月,“网络信息检索工作小组”收集整理了有关 NIR 工具及其研究团体的各种信息,编辑、出版了题为《网络信息检索水平动向:工具与团体》的调研报告。该报告介绍了 NIR - WG 所收集的各种网络信息检索工具及有关研究团体的详细信息。这是迄今为止有关网络信息检索和信息检索工具最全面也最有权威的调研报告。

二、计算机网络信息检索基本原理

计算机网络信息检索是通过计算机网络技术和设备,利用计算机网络信息检索平台,从资源平台的信息集合中识别和获取知识信息的过程。具体地讲,是检索人员根据某种目的,借助计算机网络技术和设备,在一定时间内,使用特定的检索语言和检索指令,在计算机网络中检索所需信息。

从广义上讲,计算机网络信息检索又称为信息的存贮与检索。在计算机网络中,经过描述、加工、处理与有序化的文字、资料、图像、符号等信息分别存贮在不同的计算机上,有的以数据库的形式存贮,有的以 WWW(World Wide Web)网页形式存贮,有的则以多