



面向 21 世纪 课程 教材

高等学校
信息管理类专业
核心课程教材

网络信息资源开发与利用

董慧 丁波涛 余传明 著

3

武汉大学出版社

面向 21 世纪 课程 教材

面向 21 世纪课程教材

高等学校信息管理类专业核心课教材

网络信息资源开发与利用

董慧 丁波涛 余传明 著

武汉大学出版社

图书在版编目(CIP)数据

网络信息资源开发与利用/董慧,丁波涛,余传明著 —武汉:武汉大学出版社,2001 8

面向 21 世纪课程教材.高等学校信息管理类专业核心课教材

ISBN 7-307-03184-1

I.网… II.①董… ②丁… ③余… III.计算机网络—信息管理—高等学校—教材 IV.G203

中国版本图书馆 CIP 数据核字(2001)第 22923 号

责任编辑:严红 责任校对:黄添生 版式设计:支笛

出版:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.whu.edu.cn)

发行:新华书店湖北发行所

印刷:武汉市新华印刷厂

开本:880×1230 1/32 印张:12.625 字数:334千字 插页:2

版次:2001年8月第1版 2001年8月第1次印刷

ISBN 7-307-03184-1/G·547 定价:18.00元

版权所有,不得翻印;所购我社教材,如有缺页、倒页、脱页等质量问题,请与当地教材供应部门联系调换。

内 容 简 介

本书全面介绍了当今网络信息收集、加工、整理、存储、检索、控制和利用的最先进的技术与方法；分析了利用网络信息资源、组织网络信息资源、发布网络信息资源的过程与技巧。本书是国内首部全面、系统、详细介绍网络信息资源的开发与利用的教科书，内容新颖、布局合理、重点突出、例程丰富。

本书可供高等院校信息管理类及各相关专业作教材或教学参考书，对广大科技工作者掌握网络信息资源知识和技能也很实用。

前 言

人类历史上还没有一项技术像 Internet 那样，在这么短的时间内发展速度这么快，对人类进步的影响这么大。1993 年，美国政府提出兴建“信息高速公路”这一世纪工程几个月后，日本政府就提出了“信息高速公路”计划，1994 年 2 月，欧洲也宣布兴建“信息高速公路”，接着，加拿大、韩国、新加坡、阿根廷、巴拉圭、乌拉圭等国家和我国台湾地区，也相继提出兴建“信息高速公路”计划。如此之大的工程在如此之短的时间内达到如此之统一的共识，这在人类历史上实属罕见。随之而来的发展更是令人惊叹：同是发展 5 000 万用户，电话用了 55 年，广播用了 38 年，电视用了 13 年，而 Internet 只用了不到 3 年时间；Internet 从 1993 年到 2000 年，7 年的时间已达到了 3 亿用户。在美国，到 1998 年，由网络促进其经济增长就占了 1/3 的份额。

Internet 信息资源十分丰富，是集图、文、声、像于一体的多媒体信息，其特点是：种类繁多，资源分散、无序，随机性大，信息内容特征复杂。这些特点给网络用户查询与利用带来很多困难，也对网上信息的组织与检索策略提出了新的研究课题。围绕着这个课题，作者进行了多年的努力与探索，在进行科研项目研究中，逐渐构成了本书的框架，根据 Internet 的发展，完成了本书的写作工作。本书根据信息管理与信息系统专业的特点，从信息的收集、加工、整理、存储、检索、控制到利用，吸取了当今最先进的技术、方法，意在让读者懂得如何利用网络信息资源、组织网络信息

资源、发布网络信息资源，为教学、科研及服务。

本书是教育部“高等教育面向 21 世纪教学内容和课程体系教学计划”的信息管理专业教学内容和课程体系改革成果之一，为此，十分感谢该项目组负责人康仲远教授为本书的出版创造了条件，感谢武汉大学信息管理学院院长、博士生导师马费成教授对本书出版的关心和帮助，感谢武汉大学信息管理学院博士生导师焦玉英教授对本书的关心，感谢武汉大学出版社副社长何皓先生在本书编辑、出版中给予的大力支持，感谢严红责任编辑为本书付出的辛勤劳动，感谢 98 级情报学专业硕士生杨宁、颜端武、李宏轩等同学对本书的构思提出的宝贵建议。

本教材共分九章。编写分工如下：

董 慧：第 1 章、第 2 章、第 4 章、第 8 章

丁波涛：第 5 章、第 6 章、第 7 章(除 7.4.6、7.4.7 外)

余传明：第 3 章、第 7 章(7.4.6、7.4.7)、第 9 章

全书由董慧策划、编排与统稿。限于水平和时间，加上新技术不断涌现，书中难免存在错误或不妥之处，恳请广大读者批评指正。

董 慧

2000 年 12 月于武昌珞珈山

目 录

1 Internet 基础	1
1.1 Internet 概述	1
1.2 Internet 的基本功能	16
1.3 Internet 体系结构	17
1.4 网络操作系统	18
1.5 网络数据库	33
 2 Internet 信息资源	 40
2.1 网络协议	40
2.2 域名服务	46
2.3 WWW 服务	53
2.4 电子邮件服务	55
2.5 FTP 文件传递服务	56
2.6 远程登录	57
 3 WWW	 58
3.1 浏览器	58
3.2 Internet Explorer 5	58
3.3 Netscape 浏览器	81
3.4 其它浏览器	98

4 E-mail	102
4.1 E-mail 概述	102
4.2 E-mail 工作原理	103
4.3 Outlook Express	106
4.4 电子邮件的其它功能	125
5 网页制作与网站建设	129
5.1 HTML 概述	129
5.2 HTML 的标签	129
5.3 HTML 的基本标记	134
5.4 CSS 概述	166
5.5 常用 CSS 技术	171
5.6 JavaScript 概述	186
5.7 再谈主页的制作	208
5.8 网站建设	211
6 新一代万维网的基石——XML	215
6.1 XML 介绍	215
6.2 XML 的基本规范	217
6.3 将 HTML 转换为 XML	230
7 基于 Web 的数据库开发与设计	235
7.1 概述	235
7.2 从 Client/Server 到 Internet/Web	236
7.3 Web 与数据库接口技术	239
7.4 Web 与数据库的接口方法	242
8 网络信息检索	317
8.1 概述	317
8.2 网络信息访问技术	318

8.3	网络信息检索工具的基本原理与方法	325
8.4	信息推拉技术	334
8.5	Agents 技术	341
8.6	中文 Yahoo!数据库的搜索式查询	350
9	虚拟图书馆与数字图书馆	354
9.1	虚拟图书馆概述	354
9.2	图书情报虚拟图书馆——VLLIS 系统	361
9.3	图书馆学情报学虚拟图书馆的实现	364
9.4	数字图书馆	367
附录 1	1999 年评选的中国优秀网站	373
附录 2	2000 年评选的中国优秀网站	379
附录 3	域名中英文名称对照表	383
参 考 文 献		389

1 Internet 基础

Internet 是 20 世纪对人类进步影响最大的科学贡献之一。虽然 Internet 的历史还不到半个世纪,但是它的发展速度之快,对人类进步的影响之大,却是 20 世纪其它重大科学技术所不可比拟的。回顾它的发展历史,认识其成功的重要技术因素,总结其成功的宝贵经验,可以使我们深刻地理解其技术本质,更好地开发和应用 Internet,使其更好地为人类进步服务。

1.1 Internet 概述 ■

1.1.1 Internet——20 世纪的重大科学贡献 ●

Internet 概念的形成。1961 年 MIT 的 Leonard Kleinrock 发表了第一篇介绍“分组交换理论”的学术论文,形成了今天 Internet 的理论基础。这种理论完全不同于多少年来在电话通信中采用的“电路交换理论”,而具有革命性的进步。以此为基础,从 60 年代中期开始,世界上有三个研究小组几乎同时研究基于这种“分组交换理论”的计算机网络,并且形成了基本相同的分组交换计算机网络概念。它们是 MIT(1961 ~ 1967)、RAND 公司(1962 ~ 1965)和英国国家物理实验室 NPL(1964 ~ 1967)。1968 年,美国国防部高级计划研究局 DARPA 开始资助计算机互联网的研究工作。1969 年,在美国建成了世界上第一个采用分组交换技术的计算机网络——ARPANET,将

4 所大学连接起来,基本上实现了分组交换计算机网络的构想。同年,第一个 Internet 的标准性文件 RFC 产生。

Internet 技术的实现。70 年代,世界上许多国家的通信部门看到计算机网络的巨大应用前景,纷纷开始建立提供 X.25 服务的分组交换通信网络。各计算机公司也都提出以连接本公司计算机系统为目的的计算机网络体系结构,例如 IBM 的 SNA 和 DEC 的 DNA 等。在此期间,ARPANET 的技术和规模得到了不断发展,各种各样的网络技术和应用也不断产生。人们深刻地认识到,必须设计一种标准的、开放性的计算机网络互联结构,连接各种各样的计算机网络体系结构。1974 年 Stanford 大学的 Vint Cerf 和 Bob Kahn 发表了第一篇关于分组交换网络互联协议的学术论文,提出采用传送控制协议(TCP)和网络间互联协议(IP)来实现计算机网络之间的互联。随后,DARPA 长期支持使用 TCP/IP 协议的 ARPANET 以及分组无线网和分组卫星网等一系列研究项目,从此开始了 Internet 研究的历史性进程。在此期间,美国 Hawaii 大学还在 1970 年发明了著名的无线通信网络 ALOHAnet,提出了多路共享传输介质的基本控制算法,奠定了分组无线通信和卫星通信的理论基础。1973 年,Harvard 大学的 Bob Metcalfe 在其博士论文中首次提出了 Ethernet 的基本原理,并且在同年 5 月实现了第一个 Ethernet 网络,称为 Alto Aloha 系统。至此,Internet 的基本技术已经得到实现。

Internet 功能的完善。80 年代,随着微处理器技术的迅速推广应用,以 Ethernet 为代表的计算机局域网(LAN)技术得到很大的发展。国际标准化组织 ISO 提出的开放系统互联 OSI 参考模型的研究也非常热烈。从 1983 年开始,ARPANET 正式采用 TCP/IP 协议。在此期间,产生了许多影响 Internet 发展的技术和相应的计算机网络,使 Internet 的功能不断完善。1979 年,基于 UUCP 的 Usenet 开始运行,人们在它上面开发了非常著名的 Newsgroup 功能。1981 年,建立了 BITNET,并且提供电子邮件、文件传送和 Listserver 服务。同时还发明了第一个计算机网络的 Gateway,把 BITNET 与 ARPANET 相连接并提供 E-mail 转换。同年,美国国家科学基金会(NSF)支持美

国一些大学的计算机系建立了 CSNET。1984 年,域名服务系统 DNS 开始运行,实现了使用用户域名的网络服务。1986 年,NSF 支持建立 NSFNET,形成了 Internet 的第一个主干网,当时的传输速率只有 56Kbps。在 NSFNET 上,用 IBM 的 RS6000 作为互联设备,形成了路由器的雏形。这时,Internet 的 E-mail、FTP 和 Telnet 功能已经开始实现统一的命令格式。1989 年发明了 Archie 系统,实现了网络上文件信息的查询。1991 年发明了 WAIS 系统,实现了网络上全文本文件信息的索引。1991 年发明了“客户/服务器”方式的网络应用软件体系结构,并用此结构实现了 Gopher 系统。

Internet 规模的扩展。欧洲粒子实验室的 Tim Berners-Lee 于 1991 年发明了超文本链接的 WWW 技术。Marc Andreessen 在 1993 年开发成功图形 Mosaic 浏览器系统,从此 Internet 的应用方式被更多的人接受。另外,1995 年 Sun 公司宣布了网络程序设计语言 Java,使得开发网络应用更加容易和方便。1996 年,人们发明了 Internet 电话,使 Internet 向传统电信业务和技术提出了全面挑战,于是人们对 Internet 传输速度的要求越来越高,Internet 商业化时机日渐成熟。从 1995 年开始,NSF 鼓励企业投资建设 Internet 主干网,并逐步减少对 Internet 主干网 NSFNET 的资助。由于 TCP/IP 及其应用技术的逐步发展,再加上企业资本的大量投入,Internet 以空前的速度得到发展和应用,成为当今世界信息基础设施的重要组成部分。

Internet 的技术特点。Internet 在发展和应用过程中,逐步形成了自己的主要技术特点,其中包括:

- 分组交换技术 把要传送的网络信息分成较小的分组在网络各节点上进行存储转发,成为近 30 年计算机网络通信的核心技术。

- 网络间互联协议(IP)技术 在庞大的网络 IP 地址空间中,用 IP 协议携带信息分组跨越网络中的一系列节点,独立进行路由选择,直至到达目的地。它是一种非连接的网间协议,具有非常好的联网适应能力,可以方便地互联任何类型的计算机网络,是以 IP 协议的不变去适应底层网络技术发展的万变。

- 路由器加专线连接的技术 进行路由选择的网络节点称为路由器,路由器之间的连接可以是(或可以看成是)一条通信专线。各种路由器技术和路由选择算法以及适应于实现这种通信专线功能的各种网络连接,是 Internet 技术的重要组成部分。

- 端到端的网络连接技术 用基于连接的可靠网络传送协议 TCP 和无连接的网络传送协议 UDP 为各种网络应用提供统一方便的支撑平台。

- 层次结构的网络域名管理技术 实现难以记忆的 IP 地址与具有层次结构的网络主机名之间的解析管理,使得网络的访问变得相对容易。

- 开放通用的应用技术 Internet 的基本应用技术不仅是开放的,也是通用的。

Internet 的成功经验。通过 Internet 的成功至少可以总结出以下经验:

- 长期不断的政府支持 美国政府长期支持 Internet 技术的研究达二十多年之久,终于获得了世界范围内的巨大成功,使美国的计算机网络及其应用技术领先于世界其它任何国家,同时也产生了十分巨大的经济利益。

- 具有远见的政府决策 在 Internet 发展的许多关键时刻,政府的正确决策起了至关重要的作用。例如 TCP/IP 实验网研究、NS-FNET 的建立以及 Internet 的商业化。另外,在支持 Internet 的研究过程中,给各研究单位创造公平竞争、鼓励发展的政策环境也是十分重要的因素。美国在通信基础设施建设方面先期进行了公平竞争的改革,为 Internet 的迅速商业化奠定了很好的基础。

- 技术先导的示范工程 纵观 Internet 的发展历史,我们发现,不同时期的实验研究性示范网络都建立在大学和研究单位。这样做的原因,除了因为大学和研究单位有最强的研究实力外,也因为大学是为社会培养网络及其应用人才的场所。

- 开放公开的技术标准 Internet 的技术和标准从一开始就是开放的,也是公开的。这为人们了解、参与和开发这种技术奠定了较

好的基础。

- 坚持不懈的相互合作 在 Internet 技术的发展历史中,无论是 Internet 技术研究本身,还是 Internet 的组织和管理,研究人员之间的相互合作和相互协调达到了空前的程度。

- 充满活力的企业参与 企业的积极参与和支持在 Internet 技术的发展过程中扮演了十分重要的角色。从 NSFNET 建设过程中通信公司的积极参与,到 Internet 商业化时许多企业大量投资,充当提供主干网服务的网络服务提供商(NSP)或提供 Internet 接入服务的 Internet 服务提供商(ISP),反映出现代信息产业中高风险投资的重要趋势。

1.1.2 Internet 在中国 ●

因特网在短短二十几年的时间内,已发展成为全球仅次于电话网的第二大通信网络,已从最初的教育科研网络逐步发展成为商业网络,根据 Internet Software Consortium (ISC)所作的调查,1999 年全球范围内因特网主机数增长 67%,到 1999 年年末时,已超过 7 240 万台,全球上网人数达 2.5 亿,1998 年,由网络发展而促进美国经济增长占 1/3 的份额。人类历史上还没有一项技术像因特网那样,在这么短的时间内影响如此之大。因此,因特网的发展加速了全球信息革命的进程,对发达国家,更是对发展中国家一次新的机遇与挑战。中国将在推进民族信息产业发展、加速实现国民经济信息化进程中担负起新的历史使命。

1986 年中国科学院等一些单位通过长途电话拨号方式进行国际联机数据检索,这就是我国因特网的开始。1986 年北京计算机应用技术研究所和德国卡尔斯鲁厄大学合作启动名为 CANET(Chinese Academic Network)的项目,1987 年 9 月在该所内正式建成,通过拨号 X.25 线路连通因特网电子邮件系统。CANET 作为我国第一个因特网国际电子邮件入口,向国内科研、学术、教育界提供因特网电子邮件服务,并于 1990 年 10 月正式向因特网网管中心登记注册了我国的最高域名(Top Level Domain)“CN”,从而开通了具有中国自己

域名的因特网电子邮件。

1989 年中国科学院高能物理研究所利用欧洲的计算机网关实现了 X.25 与因特网间的连接,实现了国际电子邮件转发。1990 年 CAN(Chinese Academic Network)也连通了因特网电子邮件。

1990 年中国国家计算机与网络设施 NCNFC(The National Computing Networking Facility of China)工程开工,1993 年骨干网建成。1993 年中国科学院高能物理研究所开通一条 64Kbps 的国际数据信道,和美国斯坦福线性加速中心相连。1994 年 4 月中国四大互联网之一的 CSTNET 建成,正式接入因特网。1995 年 CHINANET(中国公用网)和 CERNET(中国科研教育网)两个国家级互联网建成,同年有 100 多所高校上网。农业部信息中心通过 CHINANET 接入因特网。

中国四大互联网是:中国公用网(CHINANET)、中国金桥网(CHINAGBN)、中国科技网(CSTNet)、中国教育和科研网(CERNET)。

(1) 中国公用网(CHINANET)

CHINANET 与国内的企业网、校园网和各式各样的局域网互联,通过高速数据专线与国际因特网互联,是中国公用计算机互联和信息资源共享的骨干网,是国际因特网的重要组成部分。两年中,已实现 CHINANET 骨干网通达全国所有省会城市。目前 CHINANET 的干线速率以 2.048Mbps 为主,并将逐步升至 E3 甚至更高的速率,以适应业务发展的需要。CHINANET 在北京、上海、广州设置国际出口电路,实现与国际因特网的互联,目前总出口带宽已达 19Mbps。在 1997 年底已增开一条 45Mbps 的国际电路,以满足日益增长的用户需求。

各省现正根据本省的业务发展及用户需求情况,加快 CHINANET 各省接入层网络的建设。先期已有 20 多个省完成了其省内网的建设并开始向社会提供服务。1997 年底,其它省份的接入网的规划建设已基本完成。CHINANET 已覆盖全国 200 多个城市,基本上有业务需求的城市均已配备接入设备,方便用户的接入。CHI-

NANET 已发展成为目前国内规模最大、速率最高、用户最多的计算机互联网络。

CHINANET 除提供现有的因特网全部常规业务外,已经实现全国范围的用户漫游,使得 CHINANET 用户可以在任何省份通过 CHINANET 在当地的节点上网,使用网络资源。

因特网用户每时每刻都以惊人的速度递增,目前 CHINANET 用户每个月的增长率约为 20%,网络上的应用正在向多媒体化、宽带化方向发展,因特网的网络速率已成为网络应用发展的一个“瓶颈”,网络的能力总是赶不上用户的需要。根据业务需求规划建设好覆盖面广、高速率的基础网络,是发展好 CHINANET 的前提。

(2) 中国金桥网(CHINAGBN)

CHINAGBN 作为我国国家信息化的一项重要基础设施,经过三年多的建设,已成为国家授权的四个互联网络之一(邮电部的 CHINANET、电子工业部的 CHINAGBN、国家教育部的 CERNET、中科院的 CASNET)。CHINAGBN 是全国两家提供因特网商业服务的互联网络之一(CHINANET、CHINAGBN)。金桥工程网控中心设在北京,覆盖全国 30 个大中城市,已联网开通,初步建成了具有 PES + TES + IES + IDR 网络结构,由 X.25 分组数据交换网、综合业务数字网(ISDN)、帧中继网(Frame Relay)和计算机互联网组成的计算机综合信息通信网络,可提供 ISDN、Frame Relay、IP 等类型的网络服务。

金桥网是“天地一体”的计算机信息网络,可提供各种增值业务、多媒体业务和因特网业务,服务领域宽,综合业务能力强。金桥卫星网是目前国内技术先进、智能化程度较高的计算机信息网络。金桥网将进一步构筑 ATM 骨干网,从而实现向 B-ISDN 的过渡。

金桥网是面向社会开放的网络系统,用户包括各级政府部门、企事业单位、科研教育机构及社会大众等各种用户。金桥网将向公众提供以下服务:连接服务、网络服务、增值业务服务、多媒体信息服务、数据库联机服务等。与金桥网连接的包括国家经贸委、广电部、水利部、交通部、林业部、国家气象局、国家环保局、国家信息中心、中央电视台、中国海洋石油总公司、中国国际航空公司、中国房地产开

发集团公司、中国电子进出口总公司、广东有线电视台等单位。

(3) 中国科技网(CSTNet)

中国科技网络(China Science and Technology Network)的发展历程,实际上也就是中国因特网的发展史。

从1986年开始,国内一些科研单位,通过长途电话拨号到欧洲的一些国家,进行联机数据库检索。不久,利用这些国家与因特网的连接,进行E-mail通信。从1990年开始,北京市计算机应用研究所、中科院高能物理研究所、电子部华北计算所、电子部石家庄第54研究所等科研单位,在先后将自己的计算机与CNPAC(X.25)相连接的同时,利用欧洲国家的计算机作为网关,在X.25网与因特网之间进行转接,使得中国的CNPAC科技用户可以与因特网用户进行E-mail通信。1993年3月,中国科学院高能物理研究所(IHEP)为了支持国外科学家使用北京正负电子对撞机做高能物理实验,开通了一条64Kbps国际数据信道,连接北京西郊的中科院高能物理研究所和美国斯坦福线性加速器中心(SLAC),运行DECnet协议,虽然不能提供完全的因特网功能,但经SLAC机器的转接,可以与因特网进行E-mail通信。

正式接入因特网于1994年4月完成。中国科学院计算机网络信息中心(CNIC、CAS)自1989年10月开始,主持了“中国国家计算与网络设施”(NCFC)项目,它是世界银行贷款和国家计委共同投资的项目,内容为在中关村地区建设一个超级计算中心,供这一地区的科研用户进行科学计算。为了便于使用超级计算机,将中科院中关村地区的30多个研究所及北大、清华两所高校,全部用光缆互联在一起。其中网络部分于1993年全部完成,并于1994年3月开通了一条64Kbps的国际线路,连到美国。1994年4月,路由器开通,正式接入了因特网。

CSTNet为非盈利、公益性的网络,主要为科技用户、科技管理部门及与科技有关的政府部门服务。

中国有5000多个国有研究所,分属于中国科学院、各部委、各工业部门、各省市,已直接接入因特网的研究所只有200多个,