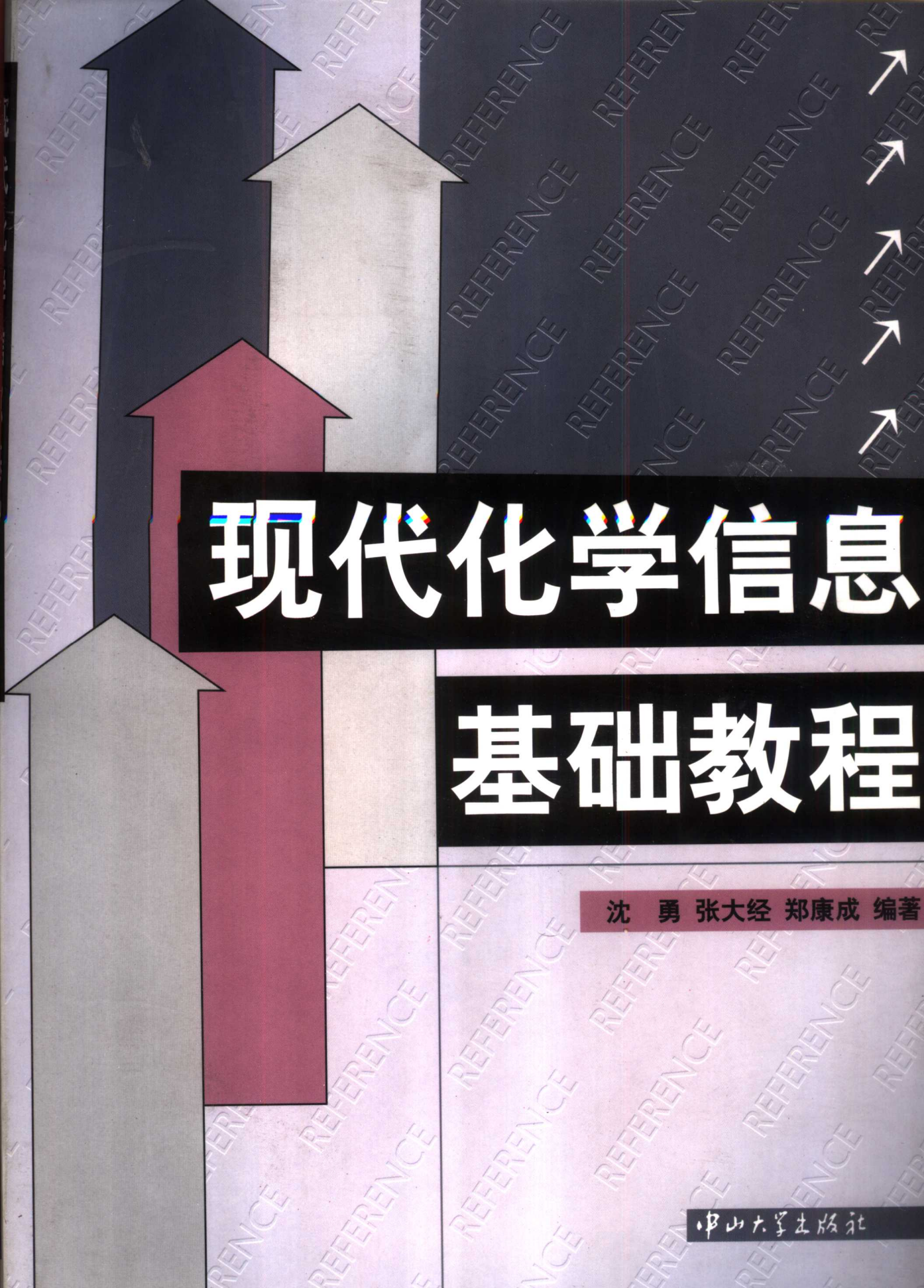




现代化学信息 基础教程

沈 勇 张大经 郑康成 编著

中山大学出版社

现代化学信息基础教程

沈 勇 张大经 郑康成 编著

中山大学出版社

·广州·

内 容 简 介

这是一本现代化学信息检索与化学图文综合处理的教科书。内容除了对现代化学文献信息与检索系统作全面介绍外,着重介绍现代化学文献的检索方法,包括网络检索、光盘检索、联机检索和手工检索等。同时介绍 ChemWin、Alchemy、HyperChem 和 ChemOffice 等现代化学图文处理软件的使用,特别介绍化学建模等方法,是化学图文处理及功能分子设计、物质结构与性质分析的重要工具,具有综合、交叉、实用与指导性的特色。

本书主要是作为化学学科本科生的教材,也适用于化工学科、生物学科的本科生,对于教师、研究生和大专生也有同样的参考和指导价值。

版权所有 翻印必究

图书在版编目(CIP)数据

现代化学信息基础教程/沈勇,张大经,郑康成编著. —广州:中山大学出版社,2000.11

ISBN 7-306-01709-8

I. 现… II. ①沈… ②张… ③郑… III. ①化学-情报检索-高等学校-教材 ②化学-信息处理-高等学校-教材 IV. G252.7

中国版本图书馆 CIP 数据核字(2000)第 49980 号

中山大学出版社出版发行

(地址:广州市新港西路 135 号 邮编:510275)

电话:020-84111998,84037215)

广东新华发行集团股份有限公司经销

番禺市市桥印刷厂印刷

(地址:番禺市市桥镇环城西路 201 号 邮编:511400 电话:84881937)

787 毫米×1092 毫米 16 开本 17.875 印张 415 千字

2000 年 11 月第 1 版 2000 年 11 月第 1 次印刷

印数:1—1000 册 定价:28.00 元

前 言

人类进入 21 世纪,现代化学研究的对象比过去大大地拓宽了。2000 年 8 月在北京召开的“首届全国化学信息学研讨会”上,徐光宪院士说,21 世纪化学的研究对象共有 10 个层次:(1)原子层次;(2)分子片层次;(3)结构单元层次;(4)分子层次;(5)超分子层次;(6)高分子层次;(7)生物分子层次;(8)介观聚集态(纳米尺度)层次;(9)宏观聚集态层次;(10)复杂分子体系层次,其中包括分子材料、分子器件(如分子导线)、分子芯片、分子机器(如分子计算机)等。如果说 19 世纪的化学是发现新元素、研究化学反应和化合物组成的技艺和科学;20 世纪的化学则是研究分子的科学;21 世纪的化学可以说是研究“广义分子”或“泛分子(Panmolecules)”的科学。

在 20 世纪 90 年代初就有人说,人类社会已进入“以情报求生存,以情报求发展”的信息时代。科学技术已成为发展生产、繁荣经济的第一生产力,并以惊人的速度创新发展。我们无意深入探讨“信息”的确切定义,但信息正在有力地推动人类社会的快速前进则是确定无疑的。当代经济学认为:

社会净产值 = 各部门投入总合 $\times$ 信息流量

这个公式指明信息及其流通在经济发展中所起的作用恰如一个“倍增器”。特别是信息高速公路的建设与发展,正在引发一场新的革命。现代信息网络作为信息社会的重要基础设施正构筑起一个崭新的信息网络环境和电子空间。信息改变着人们的生活、生产、工作和相互交往的方式。置身于信息时代的现代人才,应该学习和了解现代网络信息系统,逐步熟悉和掌握信息检索的工具与方法,以适应现代社会发展的需要。

化学化工是信息量特别大的一门学科,截至 2000 年 4 月底,在 CAS 登录的化学物质约 2400 万种。它们的各种性质,加上多元体系的性质,形成一个数据的海洋。但是,有百年历史的 Merk Index(默克索引),12thEd.(1996),收录的分子和化合物只有 10300 种,仅占 0.05%。徐光宪院士认为,总的来说,全世界的化学家用 90% 以上的精力来做研究,把成果发表在化学期刊中,但只有用不到 10% 的精力,来处理和研究化学信息宝库,总结出规律和有用结果。这好比我们挣了钱,存在银行里没有用,是很不经济的。所以,应该大力推动化学信息学的研究。

化学与其它学科相结合,不断催生一系列交叉学科和新兴学科。20 世纪 60 年代,计算机与化学结合形成了计算化学。80 年代,Internet 的飞速发展使其成为各种信息资源的重要载体,化学与信息科学的结合,逐渐成为一门崭新的化学分支学科“化学信息学”(Chemoinformatics)。

“化学信息学”首先提出于 1987 年诺贝尔化学奖获得者 J. M. Lehn 教授的获奖报告中。我们可以看出,当时化学信息的内在含义应为分子间的相互作用或相互识别的有关信息,而化学信息学应为研究分子间相互作用实质或识别机理的一门学科。“化学信息学”的定义目前尚无定论。像其它新学科一样,学科的公认或严格的定义,还需要一个研

讨、论证和完善的过程。作为一门学科应有自己的理论体系和研究的对象与目的。化学信息学用数学、统计学与计算机科学的理论方法和工具作为手段研究:(1)化学计量学;(2)化学信息处理运筹学;(3)化学模式识别,人工神经网络与定量构效关系(QSAR)的研究;(4)功能分子、材料和药物设计;(5)化学软件;(6)人工智能化工最优化工艺设计和自动控制专家系统等。

在国外一些大学已经开设了化学信息课程并确定为研究生的研究方向。在国内,化学教学指导委员会已将化学信息学列入化学教学的基本内容。北京大学化学学院、中国科学技术大学化学系等高校的院系已经开设相关课程,并编写讲义或正式出版教材。有的学校还订出本科生和研究生不同层次的化学信息学课程大纲。

本书定名为《现代化学信息基础教程》,可理解为由以前的“化学文献课”向“化学信息学”课程过渡的教材,也可以理解为本科生用的“化学信息学”的教材。本书内容涵盖化学信息表达及管理、化学文献、文献检索工具、网络化学资源、计算机网络检索方法、化学信息加工处理及化学图文处理等主要内容。本书包括 Internet 检索基础、现代化学信息检索和常用化学软件使用等三篇。

作为文献检索教材,旨在培养学生的信息意识及获取信息的能力,在基础理论和基本知识的基础上,应紧跟信息技术的发展,不断将获取信息的最新方法、手段传授给学生,不断更新教材的内容。化学化工文献检索也由过去以手工检索美国《化学文摘》和有关化学文献为主,扩大了检索的范围,增加了计算机检索等内容。随着计算机网络技术、通讯技术的高速发展,传统的文献检索迅速扩展到网络信息查询领域。

从化学信息学的角度看,它已经冲破了过去化学信息(文献)检索这一范围。在编写过程中我们力图跟上信息时代发展的步伐。但是,考虑到我国高校网络环境的实际情况和以本科生为主要对象的要求,本书既介绍化学信息检索又介绍化学图文综合处理;既介绍现代机械检索手段又介绍手工检索手段。在介绍网上信息资源时,着重介绍本科生容易获取的资源,而对国际联机检索和一些需要收费的光盘或数据库只作常识性介绍。

本书第一、三篇由沈勇执笔,第二篇由张大经执笔。郑康成负责本书的整理和审校工作。本书的出版得到本校化学学院领导和校教务处的关心、鼓励 and 大力支持。中山大学出版社在条件困难、时间紧迫的情况下给予积极的配合与支持。在编写过程中得到我校图书馆参考咨询部张颖副研究馆员和本学院刘惠兴工程师对网络检索和技术上的帮助,还得到王菊平、周豪和匡代彬等同志在化学软件翻译及图文处理上的帮助。同样,我们对上海有机所的王源教授和本书参考引用的所有文献的作者们表示深深的谢意。

本书主要作为中山大学化学学科和相关学科本科生的教材。本书的习题及上机实习指导将另出光盘。由于时间过于仓促和作者水平所限,来不及进行反复修改和验证。而且由于 Internet 化学化工资源及其相关应用正处于快速发展的时期和 Internet 资源动态性的特点,读者一定可以发现本书的一些错漏和自己从 Internet 检索时存在的某些差异。我们恳切希望专家和广大读者不吝指教,以便将来再版时加以修正与补充。

作者

2000年8月于广州

目 录

第一篇 Internet 基础

第一章 Internet 概述	(1)
第一节 计算机网络的概念	(1)
第二节 Internet 的地址管理和域名管理	(4)
第二章 Internet 的服务功能	(9)
第一节 电子邮件服务	(9)
第二节 远程登录服务	(11)
第三节 文件传输服务	(12)
第四节 Internet 新闻公告类服务	(14)
第五节 WWW 及其浏览	(16)
第三章 网络信息检索	(31)
第一节 网络信息检索概论	(31)
第二节 在 Internet 中搜索信息	(36)

第二篇 化学信息检索

第四章 信息检索基础知识	(57)
第一节 文献检索与信息检索	(57)
第二节 情报检索概述	(62)
第五章 化学文献简述	(72)
第一节 化学核心期刊	(72)
第二节 重要化学期刊简介	(75)
第三节 若干重要化学、化工工具书	(78)
第四节 《贝尔斯坦有机化学手册》	(82)
第六章 美国《化学文摘》(Chemical Abstracts)	(85)
第一节 概况	(85)
第二节 CA 的文摘	(86)
第三节 CA 的索引系统	(89)
第四节 CA 的检索步骤及检索实例	(104)
第五节 CA 的 CD-ROM 产品及使用	(106)
第七章 四大国际检索工具	(109)
第一节 《科学引文索引》	(109)

第二节 《工程索引》·····	(114)
第三节 《科技会议录索引》与《科学评论索引》·····	(120)
第八章 联机信息检索与光盘检索·····	(127)
第一节 联机信息检索和 DIALOG 系统简介·····	(127)
第二节 联机检索系统中的化学数据库·····	(129)
第三节 CISOC 化学数据库系统·····	(133)
第四节 联机检索与网络检索·····	(135)
第五节 光盘检索·····	(136)
第九章 网上信息资源与网络检索·····	(143)
第一节 搜索引擎·····	(143)
第二节 互联网上的化学信息资源与 ChIN 网页介绍·····	(147)
第三节 网上免费化学信息资源·····	(150)
第四节 网上专利信息检索·····	(166)
第五节 WWW-CAI 化学教学系统·····	(169)
第十章 信息检索技巧与实例·····	(174)
第一节 信息检索技巧·····	(174)
第二节 信息检索实例·····	(181)

第三篇 常用化学软件的使用

第十一章 ChemWindows 3.02·····	(185)
第一节 概述·····	(185)
第二节 菜单基本操作·····	(186)
第三节 基本作图方法·····	(190)
第四节 导出和导入图片·····	(190)
第五节 应用实例·····	(192)
第十二章 Alchemy III·····	(194)
第一节 Alchemy III 的简介·····	(194)
第二节 菜单命令·····	(195)
第三节 应用实例·····	(203)
第十三章 HyperChem 5.1·····	(205)
第一节 HyperChem 5.1 版本的新特性·····	(205)
第二节 菜单命令·····	(206)
第十四章 化学办公软件——ChemOffice·····	(228)
第一节 Chem3D·····	(228)
第二节 ChemDraw·····	(255)
参考文献·····	(280)

第一篇 Internet 基础

第一章 Internet 概述

随着信息时代的来临,人们获取信息的方式正发生着翻天覆地的变化。以往需要很长时间才能得到的资料,现在简单得只需轻轻点击鼠标即可完成。Internet 正是这一变化的源泉。

Internet 是当今世界上最大的信息网络,它的信息资源包罗万象,我们可以通过 Internet 实现信息查询、电子邮件收发、远程登录、远程教学等功能。要想充分利用网络资源,我们必须熟练掌握 Internet 的基础知识及相关技术。

第一节 计算机网络的概念

一、什么是计算机网络

计算机网络是指互联起来的独立的计算机的集合。要组成计算机网络,必须将两台或两台以上的计算机通过通信介质连接起来,以达到相互通信和资源共享的目的。而组成网络的计算机的工作是相互独立的,任何一台计算机的工作都不能受控于其它的计算机,任意两台机之间没有主从关系。而要真正实现计算机之间的通信,仅有物理上的连接是不够的,就像不同语言的人虽然走到了一起,但由于彼此语言不通仍无法交流。计算机之间要交流信息,也应该有统一的语言,这种语言我们称之为协议(protocol)。Internet 使用的是 TCP/IP 协议。网络协议明确地规定了数据与信息的格式和时序。只有遵循统一的网络协议,联网计算机才能够方便地交换信息、共享资源。

计算机网络是计算机技术与通讯技术紧密结合的产物。由于网络覆盖的地理范围不同,它们所采用的传输技术也不同,因而形成了不同的网络技术特点与网络服务的功能。按覆盖的地理范围进行分类,计算机网络可以分为以下三类。

(1) 局域网(LAN, Local Area Network):局域网用于将有限范围内(如一个实验室、一栋大楼、一个校园)的各种计算机、终端与外部设备互联成网。它的分布范围一般在几公里以内,最大距离不超过 10 公里。局域网具有拓扑结构简单、传输速率高、延迟小、误码率

低等特点,因此,网络站点往往能对等地参与对整个网络的使用和监控。再加上成本低、应用广、组网方便、使用灵活等特点,它深受用户欢迎,是计算机网络中最为活跃的领域之一。局域网管理软件主要有 Novell 的 Netware、Microsoft 的 Windows NT、IBM LAN Server 及 UNIX 等操作系统。

(2) 城域网(MAN, Metropolitan Area Network):城市地区网络常简称为城域网。城域网是介于广域网于局域网之间的一种高速网络。城域网设计的目标是要满足几十公里范围内的大量企业、机关、公司的多个局域网互联的需求,以实现大量用户之间的数据、语音、图形与视频等多种信息的传输功能。

(3) 广域网(WAN, Wide Area Network):广域网也称为远程网。它所覆盖的地理范围从几十公里到几千公里。广域网覆盖一个国家、地区,或横跨几个洲,形成国际性的远程网络。广域网的通讯子网可以利用公用分组交换网、卫星通信网和无线分组交换网,它将分布在不同地区的计算机系统互联起来,达到资源共享的目的。

如果说广域网的作用是扩大了信息社会中资源共享的范围,那么局域网的作用则是进一步增强了信息社会中资源共享的深度。局域网是继广域网之后网络技术与应用的又一个热点。

在 Internet 飞速发展与广泛应用的同时,高速网络的发展也引起了人们越来越多的注意。高速网技术发展主要表现在宽带综合业务数据网 B-ISDN、异步传输模式 ATM、高速局域网、交换局域网与虚拟网络上。随着技术的发展,高速局域网的数据传输速率已经达到 100 Mbps 与 1000 Mbps。

二、什么是 Internet

Internet 的中文译名是“因特网”,它是全球性的、最具影响力的计算机互联网,也是世界范围的信息资源宝库。但它本身却不是一种具体的物理网络,把它称为网络是为了便于理解而给它加上的一种虚拟“概念”。实际上它是把全世界各个地方已有的各种网络,例如计算机网、数据通信网以及公用电话交换网等互连起来,组成一个跨越国界范围的庞大的互联网。对于广大 Internet 用户来说,它好像是一个庞大的广域计算机网络。用户可以利用 Internet 实现全球范围的电子邮件、WWW 信息查询与浏览、电子新闻、文件传输、语音与图像通讯服务等功能。它对推动世界科学、文化、经济和社会的发展有着不可估量的作用。Internet 已经成为覆盖全球的信息基础设施之一。

Internet 是指全球性的信息系统,该系统具有三个重要特征:

(1) Internet 中的计算机通过全球性的惟一地址逻辑性地连接在一起。这个地址是建立在 IP(网间协议)或今后的其他协议基础上的。

(2) Internet 中的计算机之间的通信使用的是 TCP/IP 协议。

(3) Internet 可以为公共用户或个人用户提供高水平的信息服务。这种服务是建立在上述通信及相关的基础设施上的。

Internet 连接了分布在世界各地的计算机,并且按照“全球统一”的规则为每台计算机命名,制定了“全球统一”的协议约束计算机之间的交往。Internet 一开始就打破了中央控制的网络结构,任何用户都不必担心谁控制谁的问题。Internet 使世界变成了一个“地球村”,而我们每个人则变成了地球村的“村民”。

TCP/IP 协议为任何一台计算机连入 Internet 提供了技术保障。任何人、任何团体都可以加入到 Internet 中。对任何用户开放、对服务提供者开放正是 Internet 获得成功的重要原因。

从 Internet 的结构上看,它是一个使用路由器将分布在世界各地的、规模不一的计算机网络互联起来的国际网。通过 Internet,可以实现全球范围的电子邮件、WWW 信息浏览、电子新闻、文件传输、语音与图像通信服务等功能。

三、Internet 的组成部分

Internet 主要由通信线路、路由器、主机与信息资源等部分组成。

1. 通信线路

通信线路是 Internet 的基础设施,它负责将 Internet 中的路由器与主机连接起来。Internet 中的通信线路可以分为两类:有线线路(例如铜缆、光缆等)与无线线路(例如卫星与无线电等)。

我们可以使用“带宽”与“传输速率”等术语来描述通信线路的数据传输能力。所谓传输速率,是指每秒种可以传输的比特数,它的单位为位/秒(bps)。为了书写与表达方便,经常使用以下表示方法:

$$1 \text{ Kbps} = 10^3 \text{ bps} \quad 1 \text{ Mbps} = 10^6 \text{ bps} \quad 1 \text{ Gbps} = 10^9 \text{ bps}$$

通信线路的最大传输速率与它的带宽成正比。通信线路的带宽越宽,它的传输速率也就越高。

2. 路由器

路由器是 Internet 中最重要的设备之一。它负责将 Internet 中的各个局域网或广域网连接起来。当数据从一个网络传输到路由器时,它需要根据所要达到目的地为数据选择一条最佳的输出路径。如果路由器选择的输出路径比较拥挤,则由路由器负责管理数据传输的等待队列。数据从源主机出发后,往往需要经过多个路由器的转发,经过多个网络的传输才能到达目的主机。

3. 主机

主机是 Internet 中不可缺少的成员,它是信息资源与服务的载体。

按照在 Internet 中的用途,主机可以分为两类:服务器与客户机。服务器是信息资源与服务的提供者,它一般是性能比较高、存储容量比较大的计算机;而客户机是信息资源与服务的使用者,它可以是普通的微型机或便携机。服务器使用专用的服务器软件向用户提供信息资源与服务,而用户使用各类 Internet 客户端软件来访问信息资源或服务。

4. 信息资源

信息资源是用户最关心的问题,它会影响到 Internet 受欢迎的程度。Internet 的发展方向是如何更好地组织信息资源,使用户快捷地获得信息。WWW 服务的出现使信息资源的组织方式更加合理,搜索引擎的出现使信息的检索更加快捷。

在 Internet 中存在着很多类型的信息资源,例如文本、图像、声音与视频等多种信息类型,涉及到社会生活的各个方面。通过 Internet,可以查找科技资料,获得商业信息,下载流行音乐,参与联机游戏或收看网上直播等。

第二节 Internet 的地址管理和域名管理

Internet 由数十万个网络与数千万台计算机组成,而 TCP/IP 协议就是它们的纽带。虽然 Internet 的管理是松散的,但接入 Internet 的计算机必须遵循 TCP/IP 协议。TCP/IP 是一个协议集,它包括一系列的协议,并对 Internet 主机的寻址方式、命名规则、传输机制与服务功能做了详细的约定。由于 TCP/IP 协议有高可靠性、互操作性与灵活性等特点,所以能够得到广泛的应用。

一、IP 协议与 IP 地址

1. IP 协议

IP 协议即网络层协议,具有识别网络与主机号的功能,它相当于 Internet 中的交通规则。IP 协议的基本任务是 Internet 中传送 IP 数据包,各个 IP 数据包之间相互独立。

发送数据的主机使用 IP 协议将数据封装成 IP 数据包,路由器使用 IP 协议指挥 IP 数据包的传输方向,接收数据的主机使用 IP 协议将 IP 数据包拆封成数据。IP 数据包中既包括数据又包括 IP 地址信息。IP 数据包一般要经过多个网络与路由器才能到达目的地。

2. IP 地址

Internet 地址能够惟一确定 Internet 上每台计算机与每个用户的位置。Internet 地址有两种表示形式:IP 地址与域名。

接入 Internet 的计算机与接入电话网的电话相似,每台计算机或路由器都有一个由授权机构分配的号码,称为 IP 地址。IP 地址采用分层结构,它由网络号与主机号两部分组成,其结构如图 1-1 所示,长度为 32 比特,以 X.X.X.X 格式表示,每个 X 为 8 位,其值为 0~255。其中,网络号用来标识一个逻辑网络,主机号用来标识网络中的一台主机。一旦网络号和主机号的比特位数确定,就等于确定了整个 Internet 中所能包含的网络数量以及各个网络所能容纳的主机数量。Internet 上任意一台主机至少有一个 IP 地址,而且这个 IP 地址是全网惟一的。如果一台 Internet 主机有两个或多个 IP 地址,则该主机属于两个或多个逻辑网络。

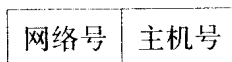
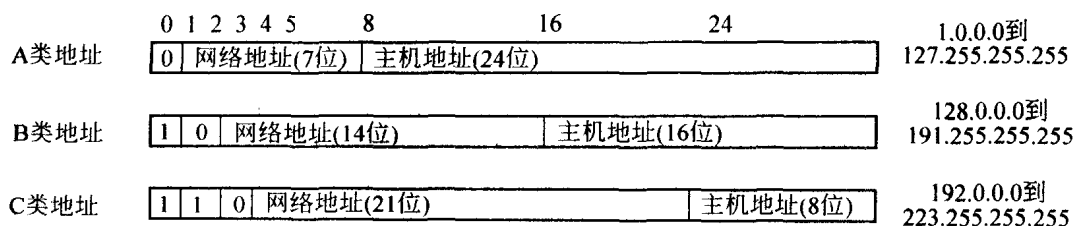
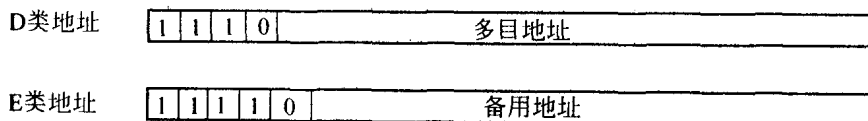


图 1-1 IP 地址的结构

在 Internet 中,网络数量是一个难以确定的因素,但是每个网络的规模却是比较容易确定的。从局域网到广域网,不同种类的网络其规模差别很大,必须加以区别。按照网络规模的大小,可以将 IP 地址分为五种类型,其中,A、B、C 是三种主要类型地址。此外,还有两种次要类型地址,一种是专供多目传送用的多目地址 D;另一种是扩展备用地址 E。IP 地址中的前 5 位用于标识 IP 地址的类别。A 类地址的第一位为“0”,B 类地址的前两位为“10”,C 类地址的前三位为“110”,D 类地址的前四位为“1110”,E 类地址的前五位为“11110”,它们的区别如图 1-2 所示。由于 IP 地址的长度限定为 32 位,所以类标识符的长度越长,可用的地址空间就越小。



(a) 主类地址



(b) 次类地址

图 1-2 基本的 IP 地址

对于 A 类 IP 地址,其网络号空间长度为 7 位,主机号空间长度为 24 位。A 类地址的表示范围为 1.0.0.0 ~ 126.255.255.255。由于网络号空间长度为 7 位,因此允许有 2^7 (126) 个不同的 A 类网络(网络号的 0 和 127 保留用于特殊目的)。同时,由于主机号空间长度为 24 位,因此每个 A 类网络的主机号数多达 2^{24} 即 16777214 个(除去主机号全为 0 和全为 1 的情况)。A 类 IP 地址结构适用于有大量主机的大型网络。

对于 B 类 IP 地址,其网络号空间长度为 14 位,主机号空间长度为 16 位。B 类 IP 地址的表示范围为 128.0.0.0 ~ 191.255.255.255。由于网络号空间长度为 14 位,因此,允许有 2^{14} (16384) 个不同的 B 类网络。同时,由于主机号空间长度 16 位,因此每个 B 类网络的主机号多达 $2^{16} - 2$ (65534) 个。B 类 IP 地址适用于有一定数量主机的中型网络。例如,清华大学网络号第一段为 162。

对于 C 类 IP 地址,其网络号空间长度为 21 位,主机号空间长度为 8 位。C 类 IP 地址的表示范围为:192.0.0.0 ~ 223.255.255.255。由于网络号空间长度为 21 位,因此允许有 2^{21} (2097152) 个不同的 C 类网络。同时,由于主机号空间长度为 8 位,因此允许每个 C 类网络的主机号数最多为 $2^8 - 2$ (254) 个。C 类 IP 地址适用于有少量主机的小型网络。

使用点分十进制编址,很容易识别出 IP 地址的类型。例如,202.113.19.122 是一个 C 类 IP 地址,其中 202.113.19.0 为网络号,122 为主机号。对于一些规模很小的局域网,即使用 C 类 IP 地址也是很浪费的,这时可以对 IP 地址中的主机号进行再次划分。例如,可以对 B 类网络号 115.108.0.0 进行再次划分,使它的第三个字节代表子网号。对于 IP 地址为 155.108.66.3 的主机,它的网络号为 155.108.66.0,主机号为 3。这时需要用子网掩码来区分 IP 地址是否经过再次划分。子网掩码也是 32 位的,网络号在子网掩码中用“1”表示,主机号在子网掩码中用“0”表示。例如,对于 IP 地址为 155.108.66.3 的主机,它的子网掩码用点分十进制表示为 255.255.255.0。

3. 几种特殊的地址

(1) 广播地址

TCP/IP 协议规定,主机号全为 1 的地址用于广播之用,称为“广播地址”。所谓广播是指同时向网上所有的主机发送报文。

(2) 有限广播地址

有时需要在本网内广播,但又不知道本网的网络号,怎么办? TCP/IP 协议规定,32 位全为 1 的地址(即“255.255.255.255”)用于本网广播。因此,该地址称为“有限广播地址”。

(3) “0”地址

TCP/IP 协议规定,各位全为“0”的网络号被解释成“本网络”。若主机试图在本网络通信,但又不知道本网的网络号,那么可以用“0”地址。

(4) 回送地址

A 类网络地址的第一段十进制数为 127 是一个保留地址,例如“127.1.1.1”,用于网络软件测试以及本地机进程间通信,叫做“回送地址”。

4. 新一代的 IP 地址

当前在 Internet 上使用的 IP 地址是在 1978 年确立的协议,当时的版本号为 4,因而称为“IPv4”,也被称为“Octet”。尽管这个协议在理论上大约有 43 亿个 IP 地址(即 2^{32})。但是,并不是所有的地址都得到充分利用,原因在于 Internet 信息中心 InterNIC 把 IP 地址分配给许多机构,而这些机构并没有充分使用所有的分配地址。例如,一些美国大学被划分为 A 类网络,每个 A 类网络所包含的有效 IP 地址为 1600 万个,这么多的地址显然没有被充分利用。由于 Internet 的迅猛发展,主机数量正在急剧增加,它正在以很快的速度耗尽目前尚未使用的 IP 地址。另外,目前剩下的地址大多属于 C 类地址,C 类网络的增加使网络数目急剧增加,将迫使 Internet 干线的路由器储存更多的网络信息,从而使网络的路由速度越来越慢。

Internet 工程工作组 IETF 提出了增加 IP 地址的两个建议:

(1) 保留 32 位格式

IETF 为了对付当前不断减少的 IP 地址,建议保留现存的 32 位地址格式 IPv4,但不再使用 A、B、C 三类划分方式,要求一些拥有 A 类或 B 类网络地址的用户放弃他们网络中尚未使用的 IP 地址,这个过程将是缓慢而困难的。

(2) 创建 IP 协议新版本——IPv6

IETF 提出的另一个建议是创建 IP 协议新版本——IPv6。IPv6 将 IP 地址扩展到 128 位,从而包含 3×10^{38} 个 IP 地址。IPv4 将以渐进方式过渡到 IPv6,IPv6 与 IPv4 可以共存。IPv6 比 IPv4 在功能方面有了许多提高:路由和寻址功能得到扩充、标题格式得到简化、增强了保密安全功能等等。IPv6 正在赢得越来越多的支持,而且很多网络硬件和软件制造商已经表示支持这个协议。然而,从 IPv4 到 IPv6 的过渡将是一个缓慢而长期的过程。

二、Internet 域名系统

Internet 中采用了统一的 IP 地址来标识主机,每台主机具有独立的 IP 地址,因而直接使用 IP 地址就可实现 Internet 中主机之间的通信。但是,由于 IP 地址是一串很难记忆的数字,所以需对应给出具有一定意义的主机名。为了提供一种直观明了的主机识别符,TCP/IP 协议专门设计了一种字符型的主机命名机制,这就是要讨论的域名系统。

1. Internet 的域名结构

Internet 的域名结构是由 TCP/IP 协议集的域名系统(DNS, DOMAIN NAME SYSTEM)来定义的。DNS 将整个 Internet 划分为多个顶级域,并为每个顶级域规定了通用的顶级域名,如表 1-1 所示。顶级域名的划分采用了两种划分模式:组织模式与地理模式。由于美国是 Internet 的发源地,因此美国的顶级域名是以组织模式来划分的。对于其他国家或地区,它们的顶级域名是以地理模式划分的,每个申请接入 Internet 的国家都可以作为一个顶级域出现。例如,cn 代表中国,jp 代表日本,fr 代表法国,uk 代表英国,ca 代表加拿大,au 代表澳大利亚。

表 1-1 顶级域名分配

顶级域名	域名类型
com	商业组织
edu	教育机构
gov	政府部门
int	国际组织
mil	军事组织
net	网络支持中心
org	各种非赢利性组织
国家或地区代码	国家或地区

网络信息中心(NIC)将顶级域的管理权授予指定的管理机构,各个管理机构再为它们所管理的域分配二级域名,并将二级域名的管理权授予其下属的管理机构。如此层层细分,就形成了 Internet 层次状的域名结构。

2. 我国的域名结构

中国互联网信息中心(CNNIC)负责管理我国的顶级域,它将 cn 域划分为多个二级域,如表 1-2 所示。我国二级域的划分采用了两种划分模式:组织模式与地理模式。其中,前七个域对应于组织模式,行政区代码对应于地理模式。例如,bj 代表北京市,sh 代表上海市,tj 代表天津市,nm 代表内蒙古自治区,hk 代表香港等。

表 1-2 二级域名分配

ac	科研机构
com	商业组织
edu	教育机构
gov	政府部门
int	国际组织
net	网络支持中心
org	各种非赢利性组织
行政区代码	我国的各个行政区

CNNIC 将 cn 域划分为多个二级域,并将 edu 域的管理权授予 CERNET 网络中心。CERNET 网络中心将 edu 域划分为多个三级域,将三级域名分配给各个大学与教育机构。例如,edu 域下的 zsu 代表中山大学,并将 zsu 域的管理权授予中山大学。中山大学又将 zsu 域划分为多个四级域,将四级域名分配给下属部门或主机。

1) 主机域名格式

Internet 主机域名的排列原则是低层的子域名在前面,而它们所属的高层域名在后面。Internet 主机域名的一般格式为:

四级域名.三级域名.二级域名.顶级域名.

在域名系统 DNS 中,每个域是由不同的组织来管理的,而这些组织又可将其子域分给其他的组织来管理。这各层次结构的优点是:各个组织在它们的内部可以自由选择域名,只要保证组织内的惟一性,而不用担心与其他组织的域名冲突。

2) 域名服务

Internet 域名系统为用户提供了极大的方便。用人们熟悉的自然语言去标识一台主机域名,自然要比用数字型的 IP 地址更容易记忆。但是,主机域名不能直接用于 TCP/IP 协议的路由选择中。当用户使用主机域名进行通信时,必须首先要将其映射成 IP 地址。这种将主机域名映射为 IP 地址的过程称为域名解析。域名解析包括两种方式:正向域名解析(从域名到 IP 地址)与反向域名解析(从 IP 地址到域名)。Internet 的域名系统 DNS 能够透明地完成此项工作。

如果要寻找一个主机名所对应的 IP 地址,则需要借助域名服务器来完成。Internet 中存在着大量的域名服务器,在每台域名服务器中保存着它所管辖区域中的主机域名与 IP 地址的对照表。当 Internet 应用程序收到一个主机域名时,它向本地域名服务器查询该主机域名对应的 IP 地址。如果在本地域名服务器中找不到该主机域名对应的 IP 地址,则本地域名服务器向其他域名服务发出请求,要求其他域名服务器协助查找,并将找到的 IP 地址返回给发出请求的应用程序。

以上讨论中涉及到这样几个术语:域、域名、域名系统、域名解析与域名服务器。域(Domain)是指某一类 Internet 主机的集合,它是管理一类 Internet 主机的一种组织形式。域名(Domain Name)是标识自然语言名称,它与数字型号的 IP 地址相对应。域名系统(Domain Name System)是管理域的命名、管理主机域名、实现主机与 IP 地址解析的系统。域名解析(Domain Name Resolving)是实现主机域名与 IP 地址映射的过程。域名服务器(Domain Name Servers)是运行域名解析软件、实现域名解析功能的 Internet 服务器。

第二章 Internet 的服务功能

Internet 的飞速发展,已使它所能提供的各种服务多达上万种,其中大多数是免费的。Internet 的商业化趋势,将会使它所能提供的服务进一步增多。本章将介绍 Internet 提供的电子邮件、文件传输、远程登录、新闻与公告类以及 WWW 等基本服务功能。

第一节 电子邮件服务

一、什么是电子邮件服务

电子邮件是随着计算机网络技术的发展而出现的一种崭新的通信手段。早在前 70 年代,美国 ARPA 的科研人员在进行 Internet 的项目研究时,为了方便科研人员之间通信,便想到利用计算机网络作为一种个人之间的通信方式。他们首先开发了使用拨号电话系统与主机相连的通信软件,不久,便诞生了应用于互连多台计算机的电子邮件系统。

二、电子邮件服务的工作过程与 SMTP 协议

1) 邮件服务器

邮件服务器是 Internet 邮件服务系统的核心,它的作用与邮局相似。一方面,邮件服务器负责接收用户送来的邮件,并根据收件人地址发送到对方的邮件服务器中;另一方面,它负责接收由其他邮件服务器发来的邮件,并根据收件人地址分发到相应的电子邮箱中。

2) 电子邮箱

要使用电子邮件服务,首先要拥有一个电子邮箱(MAILBOX)。电子邮箱是由提供电子邮件服务的机构(一般是 ISP)为用户建立的。当用户向 ISP 申请 Internet 账号时,ISP 就会在它的邮件服务上建立该用户的电子邮件账号,其中包括用户名(username)与用户密码(password)。

建立电子邮箱就是在邮件服务器的硬盘上,为用户开辟一块专用的存储空间用来存放该用户的电子邮件。用户在邮件服务器上拥有合法的账号后,就在该服务器上拥有了自己的邮箱,可以通过它来发送和接收电子邮件。任何人都可以将电子邮件发送到某个电子邮箱中,但只有电子邮箱的拥有者输入正确的用户名与用户密码后,才能查看电子邮件的内容或处理电子邮件。

3) 电子邮件地址

每个电子邮箱都有一个邮箱地址,称为电子邮件地址(E-mail address)。电子邮件地址的格式是固定的,并且在全球范围内是惟一的。

4) 电子邮件服务的工作过程

电子邮件服务基于客户机/服务器结构。首先,发送方将写好的邮件发送给自己的邮件服务器;发送方的邮件服务器接收用户送来的邮件,并根据收件人地址发送到对方的邮件服务器中;接收方的邮件服务器接收其他服务器发来的邮件,并根据收件人地址分发到相应的电子邮箱中;接收方可以在任何时间或地点从自己的邮件服务器中读取邮件,并对它们进行处理。

5) SMTP 协议

在 Internet 中电子邮件的传送都是依靠 SMTP 进行的,它的最大特点就是简单,因为它只规定了电子邮件如何在 Internet 中通过发送和接收方的 TCP 协议连接传送。而对其他操作,特别是前台的操作,例如,与用户的交互、邮件的存储、邮件系统发送邮件的时间间隔等问题均不涉及。在 Internet 中,前台与用户交互的工作是由其他程序来承担的。例如,在 UNIX 系统中,用户通过使用前台的 MAIL 和 MAILX 等程序,间接地调用了在 UNIX 系统内配置的 SENDMAIL 程序,实现 SMTP 协议对电子邮件的传送。

我们知道,计算机通信需要涉及客户机和服务器程序之间的交互。电子邮件系统也采用客户机/服务器结构。在后台,SMTP 协议就是按照客户机/服务器方式工作的。发信人的主机为客户方,收信人的邮件服务器为服务器方,双方机器上的 SMTP 协议相互配合,将电子邮件从发信方的主机传送到收信方的信箱。在传送邮件的过程中,需要使用 TCP 协议进行连接。SMTP 协议规定了发送方和接收方进行交互的动作,简单的交互模型见图 2-1。发送方的主机与接收方的邮件服务器直接相连,从而建立了从发送方主机到接收方的邮件服务器之间的直接通道,这就保证了邮件的传送十分可靠。当然,在传送邮件的过程中,双方需要交换一些应答信息,而这是通过使用 SMTP 协议的一组命令来实现的。

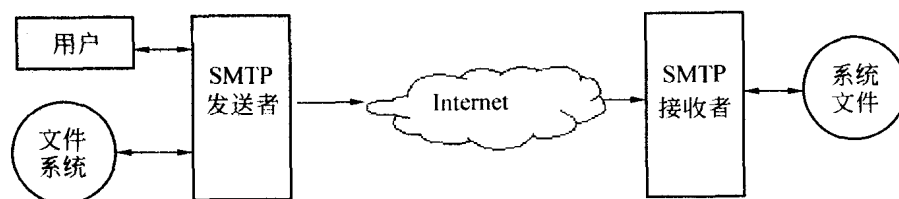


图 2-1 SMTP 协议简单交互模型

三、电子邮件应用程序

通过客户机中的电子邮件应用程序,例如 Microsoft 公司的 OUTLOOKEXPRESS,可以发送与接收电子邮件。电子邮件应用程序负责将写好的邮件发送到邮件服务器中,也负责从邮件服务器中读取邮件,并对它们进行处理。

目前,电子邮件系统几乎可以运行在任何硬件与软件平台上。各种电子邮件系统所提供的服务功能基本相同,都可以完成以下操作:

- 创建与发送电子邮件;
- 接收、阅读与管理电子邮件;
- 账号、邮箱与通信簿管理。