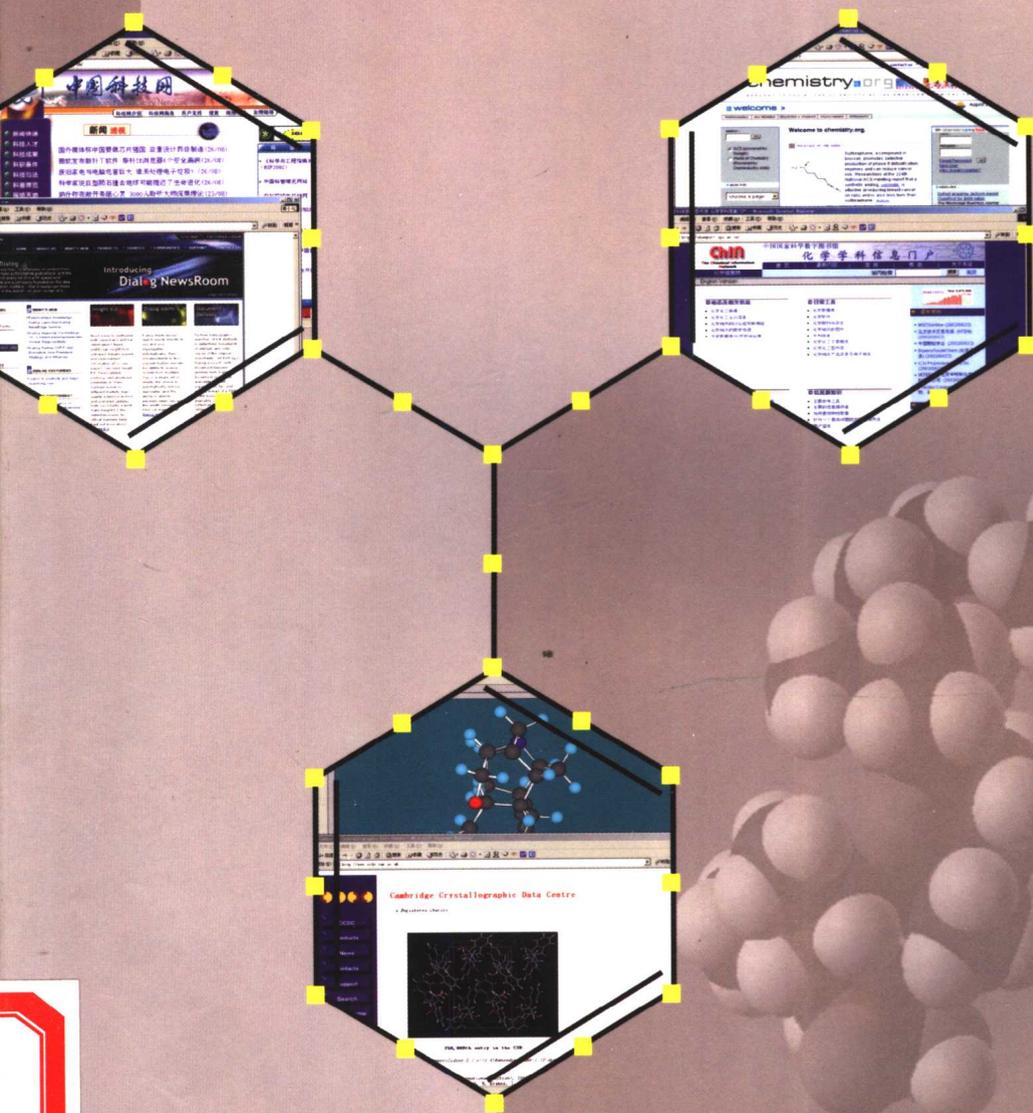


基础化学信息学

李欣 齐晶瑶 韩喜江 编著



哈尔滨工业大学出版社

基础化学信息学

李欣 齐晶瑶 韩喜江 编著

哈尔滨工业大学出版社
·哈尔滨·

内容简介

本书是一本介绍化学信息学及其应用的技术型图书。以实用性为特点,强调实践并侧重于计算机技术与计算机网络技术的应用。全书主要内容为:Internet 上的化学信息检索工具,化学数据库、化学虚拟图书馆与电子期刊,分子图形学软件,计算化学软件与资源,化学专家系统,计算机辅助合成设计,化学 CAI 教育与远程教育信息资源,虚拟仪器及其在化学中应用,科学计算可视化,Internet 与化学软件等。书中介绍了很多常见的、功能强大的化学软件与系统,如 ISIS/Draw, ChemWindow, ChemSketch, ChemOffice, DS ViewerPro, HyperChem, KnowItAll, ChemLab, DENDRAL, LHASA 等等。

本书可作为高等院校化学化工及其相关专业的“化学信息学”教材,亦可作为相关领域的广大科研工作者的参考书和应用工具书。

图书在版编目(CIP)数据

基础化学信息学/李欣编著. —哈尔滨:哈尔滨工业大学出版社,2003.3

ISBN 7-5603-1831-2

I.基… II.李… III.计算机应用-化学-情报检索 IV.G252.7

中国版本图书馆 CIP 数据核字(2003)第 012384 号

出版发行 哈尔滨工业大学出版社
社 址 哈尔滨市南岗区教化街 21 号
邮 编 150006
传 真 0451-6414749
印 刷 肇东粮食印刷厂
开 本 开本 787×1092 1/16 印张 15.5 字数 350 千字
版 次 2003 年 3 月第 1 版 2003 年 3 月第 1 次印刷
书 号 ISBN 7-5603-1831-2/O·141
印 数 1~3 000
定 价 19.00 元

前 言

21 世纪是计算机世纪,是信息世纪。化学信息学(Chemoinformatics)就是伴随着计算机技术和计算机网络技术的飞速发展而形成的一个新的化学分支,它是以化学、化工、计算机与信息技术为基础的一门边缘学科。

培养学生信息的获取和利用的能力,在世界各国的高等教育中都占有重要的地位。国外发达国家有很多大学已经开设了化学信息学类课程,1998 年 10 月我国教育部理科化学教学指导委员会也把化学信息学列入了高等学校化学专业和应用化学专业的教学基本内容之中,中山大学化学化工学院、中国科技大学化学系等高校已陆续开设相关课程。

本书的重点不是诠释计算机应用理论和分析软件结构,而是强调实践并侧重于计算机与网络的应用。主要内容为 Internet 上的化学信息检索工具,Internet 上的化学数据库,Internet 上的化学虚拟图书馆与电子期刊、分子图形学软件、计算化学软件与资源、化学专家系统、计算机辅助合成设计、化学 CAI 教育与远程信息资源、虚拟仪器及在化学中应用、科学计算可视化,以及 Internet 与化学软件。书中介绍了很多常见的、功能强大的化学软件与系统,如 ISIS/Draw, ChemWindow, ChemSketch, ChemOffice, DS ViewerPro, Hyperchem, CAVE, KnowItAll, ChemLab, DENDRAL, LHASA 等等。

本书曾作为校内讲义在哈尔滨工业大学化学及相关学科的本科和研究生教学中试用。在本书的编写过程中,参考了大量的 Internet 上的资源及相关参考书,由于资源信息的庞杂,我们无法向各种信息源的所有者一一发函征求意见,在此对所有提供帮助的网站和作者一并表示衷心的感谢。由于作者水平有限,书中的疏漏之处在所难免,敬请读者批评指正。

本书可作为高等学校化学专业、应用化学专业、环境科学专业、环境化工专业及相关专业教材,亦可供同类专业技术人员参考。

作者

2002 年 10 月

于哈尔滨工业大学

目 录

第 1 章	INTERNET 上的化学信息检索工具	(1)
1.1	Internet	(1)
1.2	INTERNET 上的通用检索系统及工具	(8)
1.3	化学专业型资源检索引擎	(12)
第 2 章	INTERNET 上的化学数据库	(18)
2.1	化学专业型数据库	(18)
2.2	专利型数据库	(26)
2.3	通用数据库	(33)
第 3 章	INTERNET 上的化学虚拟图书馆及电子期刊	(50)
3.1	虚拟图书馆	(50)
3.2	电子期刊	(52)
第 4 章	分子图形学软件	(65)
4.1	分子图形学简介	(65)
4.2	ISIS/Draw 软件	(71)
4.3	Chem Window 软件	(76)
4.4	ACD/Chemsketch 软件	(78)
4.5	ChemDraw Pro 软件	(90)
4.6	DS ViewerPro 软件	(94)
4.7	其他软件简介	(100)
4.8	分子图形学中的虚拟现实技术	(103)
第 5 章	计算化学软件与资源	(112)
5.1	计算化学简介	(112)
5.2	分子模拟简介	(114)
5.3	HyperChem 软件	(116)
5.4	Chem3D 软件	(118)
5.5	其他计算化学软件简介	(122)
5.6	远程计算服务	(123)

第 6 章 化学专家系统	(130)
6.1 人工智能及专家系统简介	(130)
6.2 化学专家系统	(131)
6.3 KnowItAll 信息科学系统	(137)
6.4 ChemLab 的结构与谱图解析专家系统	(144)
第 7 章 计算机辅助合成设计	(146)
7.1 有机合成简介	(146)
7.2 计算机辅助合成设计	(148)
7.3 分子设计	(154)
7.4 组合化学与绿色化学	(160)
第 8 章 计算机辅助化学教学	(168)
8.1 计算机辅助化学教学概述	(168)
8.2 计算机化学教学软件	(169)
8.3 化学实验教学软件	(170)
8.4 仪器分析仿真软件	(173)
8.5 网上化学教学资源	(181)
第 9 章 虚拟仪器及其在化学中的应用	(185)
9.1 虚拟仪器简介	(185)
9.2 虚拟仪器的编程语言 LabView	(188)
9.3 虚拟仪器的编程语言 LabWindows/CVI	(195)
9.4 虚拟仪器在化学中的应用	(201)
第 10 章 科学数据处理软件及可视化	(207)
10.1 MATLAB 软件	(207)
10.2 Origin 软件	(222)
10.3 科学计算可视化	(225)
第 11 章 INTERNET 对化学软件的影响	(230)
11.1 CambridgeSoft 软件	(230)
11.2 MDL 软件	(231)
11.3 Accelrys 软件	(232)
11.4 ACD 软件	(233)
11.5 ChemSW 软件	(234)

11.6 化学工程进展 CEP 的软件目录.....	(235)
11.7 “SoftwareReviews”.....	(235)
参考文献.....	(237)

第 1 章 INTERNET 上的化学信息检索工具

1.1 Internet

1.1.1 Internet 简介

在网络出现之前,用软盘传递数据,是计算机之间数据共享的惟一方式(图 1.1)。互连网的出现,不仅能够实现计算机资源的共享,而且由于它的强大功能,通过 Internet 可以在全球范围内查找你感兴趣的任何信息。

Internet(Interconnect + Networks)是一个由各种不同类型和规模的大小计算机网所组成的全球性的计算机网络(图 1.2)。Internet 的出现可以说是美国与前苏联冷战的产

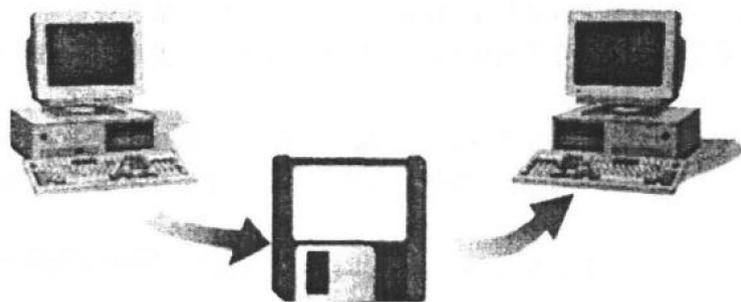


图 1.1 网络出现之前的数据共享方式

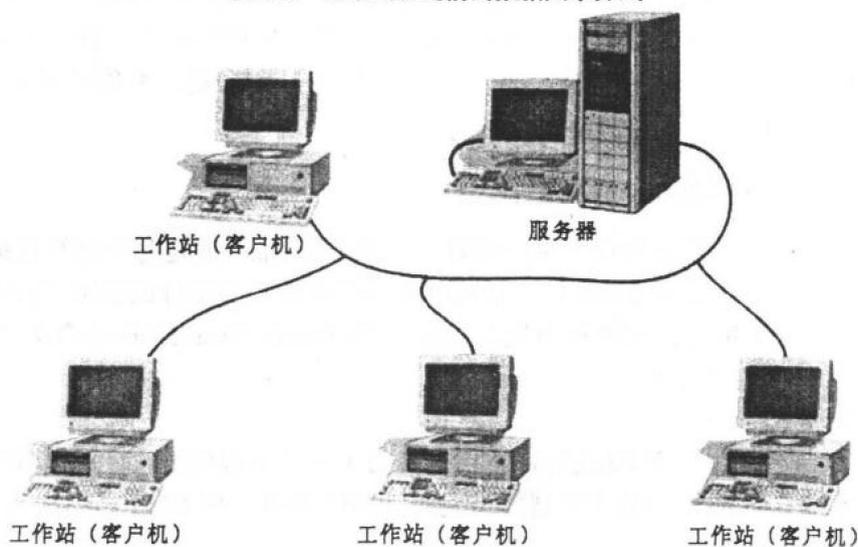


图 1.2 包含文件服务器的局域网

物。1969年末,美国国防部出于军事目的建立了 ARPAnet (Advanced Research Projects Agency Network 高级研究计划局)项目。ARPAnet 最初设计因特网的目的是用以支持军队的一项研究,即设计建立一个网,当它的一部分被迫停止运行时(例如遭核弹袭击),此网仍能完成工作。20世纪90年代,Internet 向公众开放,并迅速成为世界上最大的全球性信息网络。

Internet 的优点:(1)Internet 是一个开放的网络,不为某个人或某个组织所控制,人人都可自由参与;(2)信息量大,内容丰富;(3)不受时间、空间的限制;(4)入网方便,操作简单;(5)可以迅速便宜地实现通信和信息资源交换。

1.1.2 TCP/IP 协议简述

TCP/IP 协议是 Internet 中最为重要的关键技术之一。TCP/IP 是上百个(用来连接计算机和网络的)协议合起来的共有名字。TCP/IP 的实际名字是来自最重要的两个协议,TCP (Transfer Control Protocol 传输控制协议)和 IP(Internet Protocol 互连网协议)。

在 Internet 内部,信息不是一个恒定的数据流从一台计算机传送到另一台计算机,而是把数据分解成小包,即数据包。例如,当你传送一个很长的信息给远方的朋友时,TCP 就把这个信息分成很多个数据包,每一个数据包用一序号和一接收地址来标定。此外,为了使接受方知道数据在传输过程中是否出错,TCP 还插入一些校验信息。

把数据包通过网络传输,这就是 IP 的工作,即把它们传送给远程主机。在 Internet 上每台主机都有一个特定的地址(称为 IP 地址)。IP 的任务就是根据地址,在两台计算机之间寻找一条适当的路由。

在接受端的主机上,TCP 接收到数据包并检查错误。如果有错误发生,TCP 可以要求重发这个特定的数据包。当全部数据包都被正确地收到后,TCP 再根据序号重新组合成原来的信息。

简言之,IP 的工作是把原始数据从一地传送到另一地;TCP 管理这种流动并确保其数据正确。

除了 TCP 和 IP 协议外,TCP/IP 集中还有其他协议,例如文件传输协议 FTP(File Transfer Protocol),点对点通信协议 PPP(Point to Point Protocol),用户数据包协议 UDP(User's Datagram Protocol),简单邮件传输协议 SMTP(Simple Mail Transfer Protocol),超文本传输协议 HTTP(Hypertext Transfer Protocol)等。

1.1.3 Internet 地址与域名简介

Internet 上的每台计算机和每个用户都有一个地址,能惟一确定每台计算机的位置,区分每个不同用户。通常看到的地址是一组词语或有隐含意义的字母或数字,它们均按标准格式用“.”隔开,这种地址的编排称为域名系统(DNS-Domain Name System)编址。地址可以写成 IP 地址和域名两种形式。

1. IP 地址格式

每个 IP 地址都由 4 个字节组成,每个字节相当于一个 8 位位组,每个 8 位位组的二进制数可以表示成 0~255 之间的十进制数,数字之间用点隔开。例如“210.46.66.194”就表示某台计算机的 IP 地址。

2. 域名格式

Internet 域名是单位或个人在 Internet 网上的名称和地址,它是一种网络概念上的“商标”。域名的形式是以若干个英文字母、数字及连接符“-”组成,由“.”分隔成几个部分。

其中,顶级域(TLD)是指域名层次的最高一级,二级域(SLD)是紧跟在顶级域下的那一级域。通用顶级域(g TLD)是指“.com”、“.org”和“.net”顶级域及在“Internet 登记者委员会谅解备忘录”中或授权下建立的顶级域。

顶级域名定义为三类:

国家顶级域(n TLD):国家顶级域名的代码由 ISO3166 规定,比如.cn 代表中国,.jp 代表日本。国际顶级域(i TLD):.int,国际联盟、国际组织可以在.int 下注册,比如世界知识产权组织的域名为:wipo.int。

通用顶级域(g TLD):根据 1994 年 3 月公布的 RFC1 591 规定,通用顶级域是:

- . com 公司、银行、厂矿、新闻出版等企业单位;
- . net 网络服务机构
- . org 非盈利性组织
- . edu 教育机构
- . gov 政府部门
- . mil 军事部门

由于历史原因,IAHC 认为 . edu、. gov、. mil 是特殊域名,作为美国专用。另外,IAHC 又新增加了七个通用顶级域名,分别为:

- . firm 公司企业
- . store 销售公司企业
- . web 突出万维网(www)活动实体
- . arts 主要从事文化娱乐活动的实体
- . rec 主要从事消遣娱乐活动的实体
- . info 主要从事信息服务的实体
- . nom 一些希望在互联网上发布个人信息的人。

中国互联网络 CN 域下的二级域分为“类别域”和“行政区域”。

类别域:

- . ac. cn 科研机构
- . com. cn 工、商、金融等企业
- . edu. cn 教育部门
- . gov. cn 政府部门
- . net. cn 互联网络,接入网络的信息中心和运行中心
- . org. cn 各种非盈利性的组织。

行政区域:

有 3 4 个,适用于我国的各省、自治区、直辖市,分别为:BJ 北京市,SH 上海市,TJ 天津市,CQ 重庆市,HE 河北省,SX 山西省,NM 内蒙古自治区,LN 辽宁省, JL 吉林省,HL 黑龙

江省, JS 江苏省, ZJ 浙江省, AH 安徽省, FJ 福建省, JX 江西省, SD 山东省, HA 河南省, HB 湖北省, HN 湖南省, GD 广东省, GX 广西壮族自治区, HI 海南省, SC 四川省, GZ 贵州省, YN 云南省, XZ 西藏自治区, SN 陕西省, GS 甘肃省, QH 青海省, NX 宁夏回族自治区, XJ 新疆维吾尔自治区, TW 台湾, HK 香港, MO 澳门。

1.1.4 Internet 的功能

1. 基本服务

(1) 电子邮件(E-mail)。一种利用网络交换信息的非交互式服务。可以收发、转发信件,也可以同时向多个用户发送信件。电子邮件的内容可以是文本文件,也可以是影像、声音等多媒体信息。

(2) 文件传输服务(FTP)。用于直接进行文件双向传输的服务功能,下载(download), 上传(upload)。

(3) 远程登录(Telnet)。使用户通过网络成为远程计算机的终端用户,实现资源共享。

2. 公告板服务(BBS)

利用电子通信手段,“张贴”各种公告信息或消息。

3. 交互式信息传递服务

(1) 信息查询工具(Gopher)。基于菜单驱动的信息查询软件,它可以自动将用户的请求转换为 FTP 或 Telnet 命令,用户在菜单引导下可对 Telnet 上的远程信息系统进行联机访问。

(2) 全球信息网(WWW)。World Wide Web(WWW)目前是 Internet 上发展最快的服务,是一种多媒体信息查询工具。除上述服务功能外,Internet 还提供 Archie、广域信息服务(Wais)、电子新闻(USENET News)、电子论坛(Listserv)、在线交谈、网络会议(Netmeeting)等多种服务。

1.1.5 Internet 在中国

目前,我国有四个国际出口通向 Internet 网。

1. 中国科学院的中国科技网(CASNET)

中国科技网(CSTNET, <http://www.cnc.ac.cn/>, 图 1.3)是在北京中关村地区教育与科研示范网和中国科学院计算机网络的基础上建设和发展起来的覆盖全国范围的大型计算机网络,是我国最早建设并获国家承认的具有国际信道出口的中国四大互联网络之一。

中国科技网为非盈利、公益性的网络,也是国家知识创新工程的基础设施,主要为科技界、科技管理部门、政府部门和高新技术企业服务。

中国科技网始建于 1989 年,并于 1994 年 4 月首次实现了我国与国际互联网络的直接连接,同时在国内开始管理和运行中国顶级域名 CN。

中国科技网现有多条国际信道连到美国及日本,进入 Internet 国际互联网络。1999 年 12 月 CSTNET 国际卫星信道升级至 10Mbit。2000 年 11 月 CSTNET 国际出口带宽升级至 55Mbit。2000 年 4 月 CSTNET 以 155Mbit 接入国家互联网交换中心北京点,完成与国内其他九大互联网的互连互通。2001 年 6 月 CSTNET 开通与 CHINANET 的 155Mbit 直连信道。目

前,中国科技网在全国范围内已接入农业、林业、医学、地震、气象、铁道、电力、电子、航空航天、环境保护和国家自然科学基金委员会、国家专利局、国家计委信息中心、高新技术企业,以及中国科学院分布在北京地区和全国各地 45 个城市的科研机构,共 1 000 多家科研院所、科技部门和高新技术企业,上网用户达 40 万人。

中国科技网的服务主要有网络通信服务、域名注册服务、信息资源服务和超级计算服务。

中国科技网上的科技信息资源有科学数据库、中国科普博览、科技成果、科技管理、技术资料、农业资源和文献情报等。其中科学数据库现有专业数据库 180 个,总数据量达 725Gbit,可以向国内外用户提供各种科技信息服务。

中国科技网的超级计算中心拥有数台超级计算机,总计算能力在浮点运算每秒 1 300 亿次以上,可以通过网络向全国科技和教育人员提供高性能科学计算服务。

中国科学院计算机网络信息中心是中国科技网的网络管理运行中心。中国科学院计算机网络信息中心经国家主管部门授权,管理和运行中国互联网络信息中心,并向全国提供网络域名注册服务。

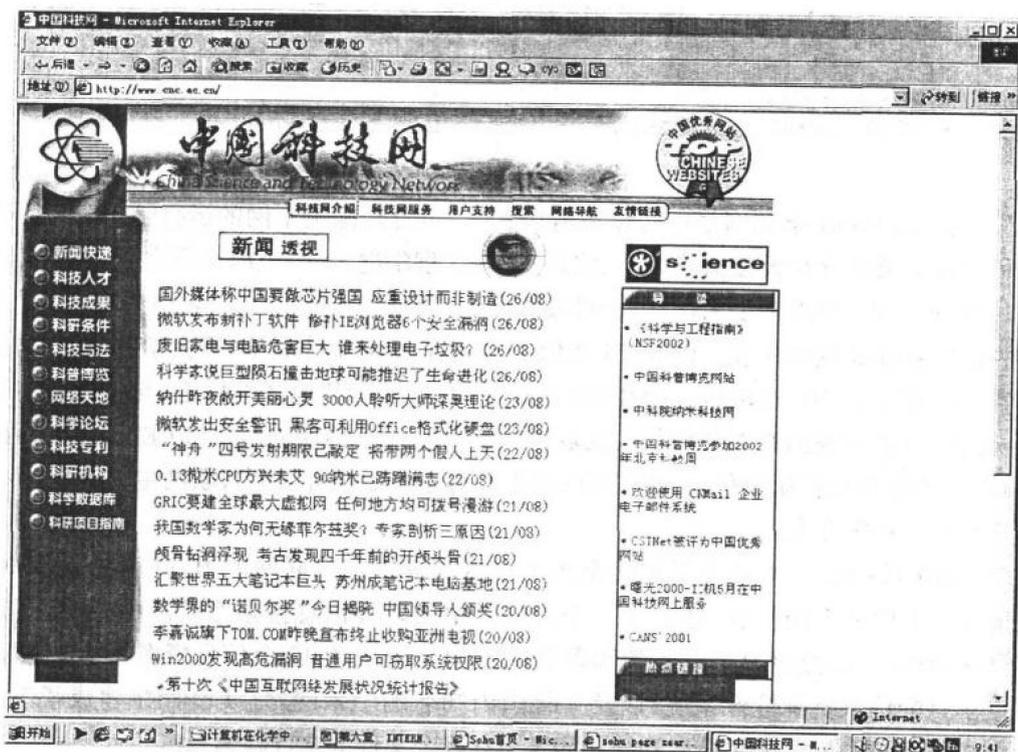


图 1.3 CASNET 主页

2. 国家教育部的中国教育与科研计算机(CERNET)网

中国教育和科研计算机网 CERNET(<http://www.edu.cn/>,图 1.4)是由国家投资建设,教育部负责管理,清华大学等高等学校承担建设和管理运行的全国性学术计算机互联网络。它主要面向教育和科研单位,是全国最大的公益性互联网络。

CERNET 分四级管理,分别是全国网络中心,地区网络中心和地区主结点,省教育科研



图 1.4 CERNET 主页

网,校园网。CERNET 全国网络中心设在清华大学,负责全国主干网的运行管理。地区网络中心和地区主结点分别设在清华大学、北京大学、北京邮电大学、上海交通大学、西安交通大学、华中科技大学、华南理工大学、电子科技大学、东南大学、东北大学等 10 所高校,负责地区网的运行管理和规划建设。CERNET 省级结点设在 36 个城市 的 38 所大学,分布于全国除台湾省外的所有省、市、自治区。CERNET 已经有 28 条国际和地区性信道,与美国、加拿大、英国、德国、日本和我国香港特别行政区联网,总带宽达到 250Mbit/s。与 CERNET 联网的大学、中小学等教育和科研单位达 900 多家(其中高等学校 800 所以上),联网主机 120 万台,个人用户达到 800 多万人。

CERNET 目前已基本具备了连接全国大多数高等学校的联网能力,并完成了 CERNET 八大地区主干网的升级扩容,建成了—个大型的中国教育信息搜索系统。CERNET 建成了总容量达 800Gbit 的全世界主要大学和著名国际学术组织的 10 个信息资源镜像系统和 12 个重点学科的信息资源镜像系统,以及—批国内知名的学术网站。CERNET 建成了系统容量为 150 万页的中英文全文检索系统和涵盖 100 万个文件的文件检索系统。

CERNET 是我国开展现代远程教育的重要平台。为了适应国家“面向 21 世纪教育振兴行动计划”中远程教育工程的要求,1999 年,CERNET 开始建设自己的高速主干网。利用国家现有光纤资源,在国家 and 地方共同投入下,到 2001 年底,CERNET 已经建成 20 000 km 的 DWDM/SDH 高速传输网,覆盖我国近 30 个主要城市,主干总容量可达 40Gbit/s;在此基础上,CERNET 高速主干网已经升级到 2.5Gbit/s,155Mbit 的 CERNET 中高速地区网已经连接到我国 35 个重点城市;全国已经有 100 多所高校的校园网以 100 ~ 1 000Mbit 速率接入 CER-

NET。

CERNET是中国开展下一代互联网研究的试验网络,它以现有的网络设施和技术力量为依托,建立了全国规模的 IPV6 试验床。1998 年 CERNET 正式参加下一代 IP 协议(IPV6) 试验网 6BONE,同年 11 月成为其骨干网成员。CERNET 在全国第一个实现了与国际下一代高速网 INTERNET 2 的互联,目前国内仅有 CERNET 的用户可以顺利地直接访问 INTERNET 2。

3. 中国网络通信公司的中国公共计算机互联网(ChinaNET)

中国公共计算机互联网(ChinaNET, <http://www.chinanetcom.com.cn/>, 图 1.5)按自然地理区域分为北京、上海、华北等 8 个大区,构成 8 个核心层节点,围绕 8 个核心节点形成 8 个区域,共 31 个节点。网络结构分为全国骨干网、各省市地区网以及接入网 3 个层次。该网现有北京及上海两个国际出入口。



图 1.5 ChinaNET 主页

4. 原电子部的金桥网

金桥网(<http://www.gb.com.cn/>, 图 1.6)是覆盖全国的由卫星网和光纤网构成的天地一体化网络结构。金桥网络初步形成了全国骨干网、省网、城域网 3 层网络结构,其中骨干网和城域网已初具规模,覆盖城市超过 100 个。网络采用了 DWDM、SDH、ATM、IP 等多种先进的技术,全程全网统一管理,具有开放式网络构架,调度灵活,可以承载包括语音、数据和图像等多种综合业务。据统计,截止到 2001 年底,金桥网络在全国的 100 多个城市建成区域汇节点和节点,拥有北京、上海、广州等 3 个国际出口,实现了世界范围内的互联互通,做到语音与数据的融合;同时在北京建立网络控制中心,实行全国统一管理,由上海的网络控

制中心提供部分备份;建成了数百个 VAST 卫星小站。

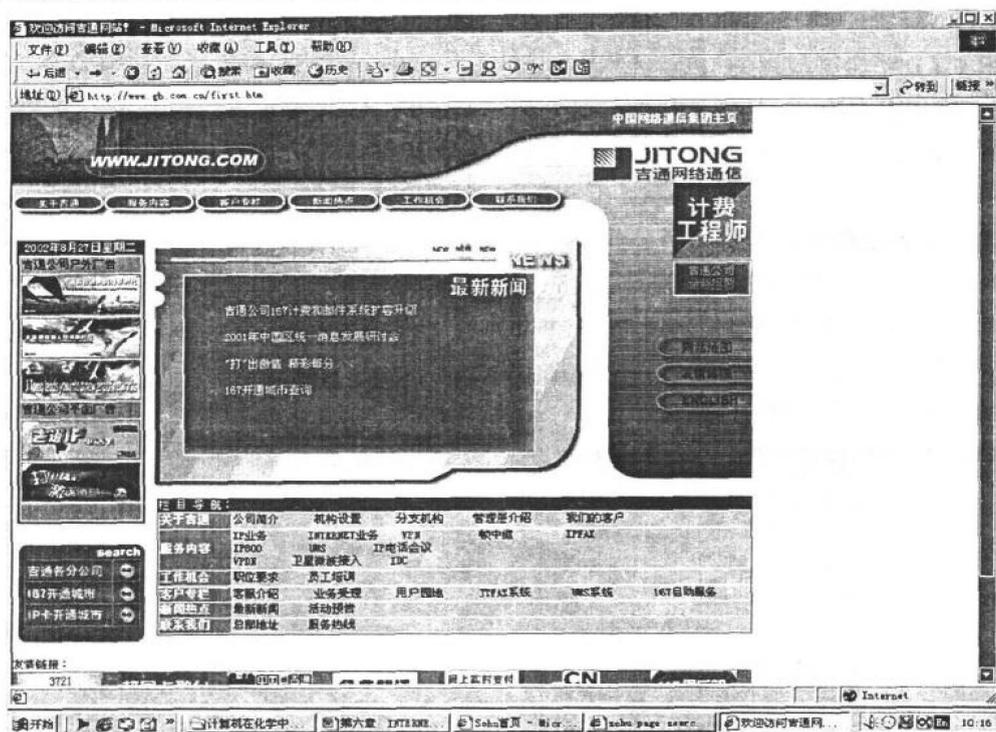


图 1.6 金桥网主页

1.2 INTERNET 上的通用检索系统及工具

1.2.1 国际联机检索系统(On-Line Retrieval)

联机检索,是指用户通过终端和通信线路,从信息检索系统存储的文献数据库中检索自己所需的信息。国际联机检索系统是以大型计算机为核心,大容量存储设备为基础,以公共数据交换网或电话通信线路为传输媒介,建立起来的大型计算机情报检索系统。

1. DIALOG 系统(<http://www.dialog.com>)

DIALOG 系统是国际上最著名、规模最大的联机检索系统。目前,该系统拥有三台大型计算机,总存储量达 180Gbit,与两个卫星通信网络相连,在全世界 80 多个国家和地区有 10 多万个终端用户,系统每天 24 小时为用户提供联机检索服务,图 1.7 为 DIALOG 主页。目前该系统建有 400 多个文档,文献存储量达 2 亿多个记录,文档的专业范围遍及科学技术、政治经济、商业、新闻等领域。DIALOG 公司还将数据库文档制作成各专业的数据库光盘,单独出版。

2. BDS 系统(<http://bds.cetin.net.cn/>)

BDS 系统是我国最大的联机检索系统,属北京文献服务处(Beijing Document Service,简称 BDS),图 1.8 为 BDS 主页。北京文献服务处是 1978 年由中国国防科技信息中心和北京市科协共同策划联合组建的,以联机信息检索服务为主,同时进行信息技术应用研究开发的

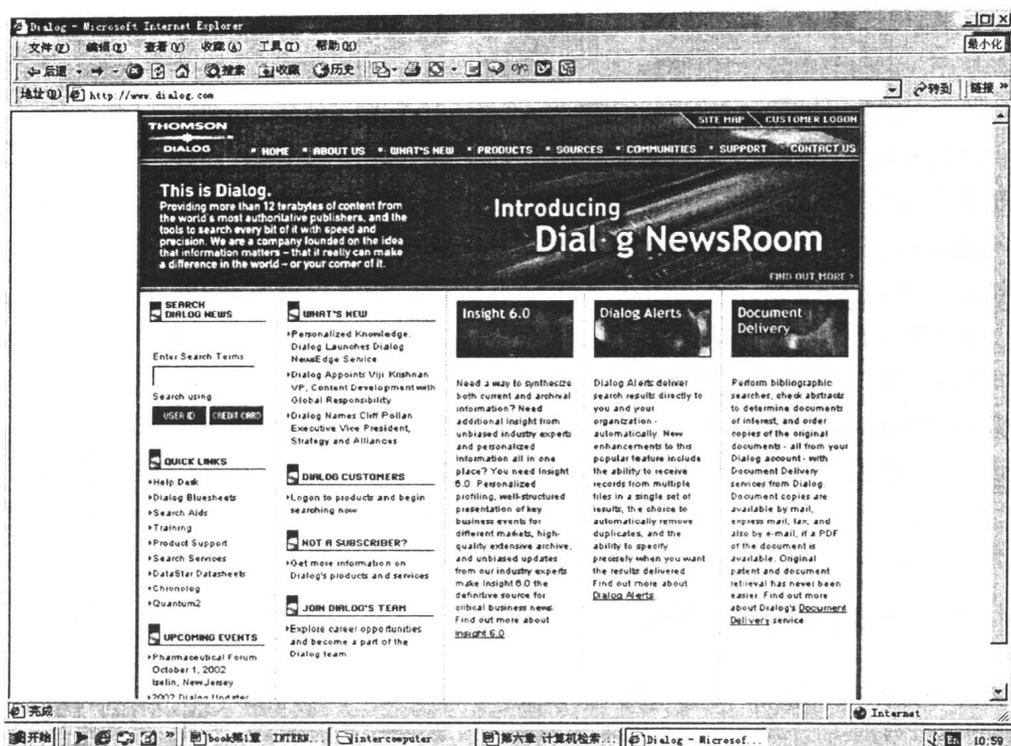


图 1.7 DIALOG 主页

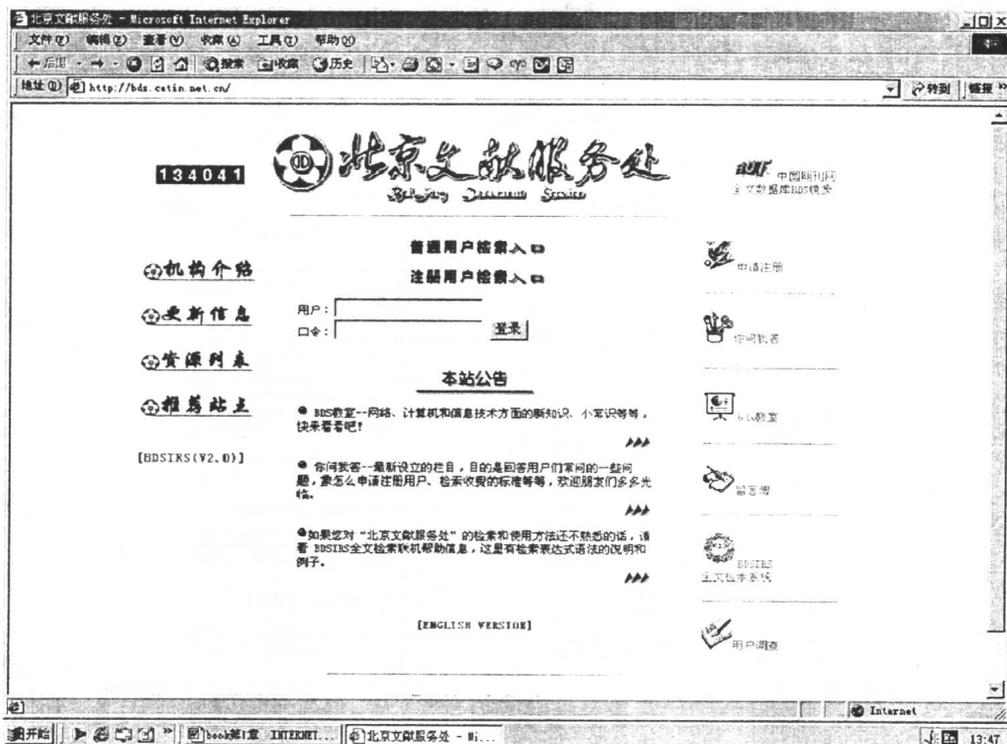


图 1.8 BDS 主页

综合性机构。

BDS 于 1981 年建立了中国第一个计算机联机信息检索系统,经过十几年的发展,已经成为国内系统配置最大、信息量最多的联机信息服务中心。

BDS 拥有强大的网络互联和信息处理能力。由国防科工委和其他工业部联合投资,筹建了覆盖全国的中国工程技术信息网, BDS 是这个大型网络的管理控制、信息服务中心,负责全网的运行、管理、信息资源集成和服务;同时开展信息处理技术、以信息检索为主的信息应用技术的研究和开发。在大型服务器上自行开发的全文检索系统,为用户提供了 40 多种,2 200 万篇文献数据库的联机检索服务。光盘服务系统提供了十几种近 300Gbit 的多媒体光盘数据库的联网服务。除提供信息检索服务外, BDS 还向用户提供方便可靠的网络信息服务和 INTERNET 互联服务。

该系统的数据库主要是专利文献和美国政府报告。该系统在全国各地拥有数百个终端用户,由于国内的联机检索费用较低,因而利用率要高于国际联机检索。但国内联机检索数据量太少无法满足各专业的用户需求。

1.2.2 通用型检索引擎

1. Yahoo(<http://www.yahoo.com>)

Yahoo 是目前 INTERNET 上规模最大、最受欢迎的搜索引擎之一,主页如图 1.9 所示。Yahoo 最成功的地方是它利用了约 150 人的编辑队伍建立了最大的 Web 资源目录系统。

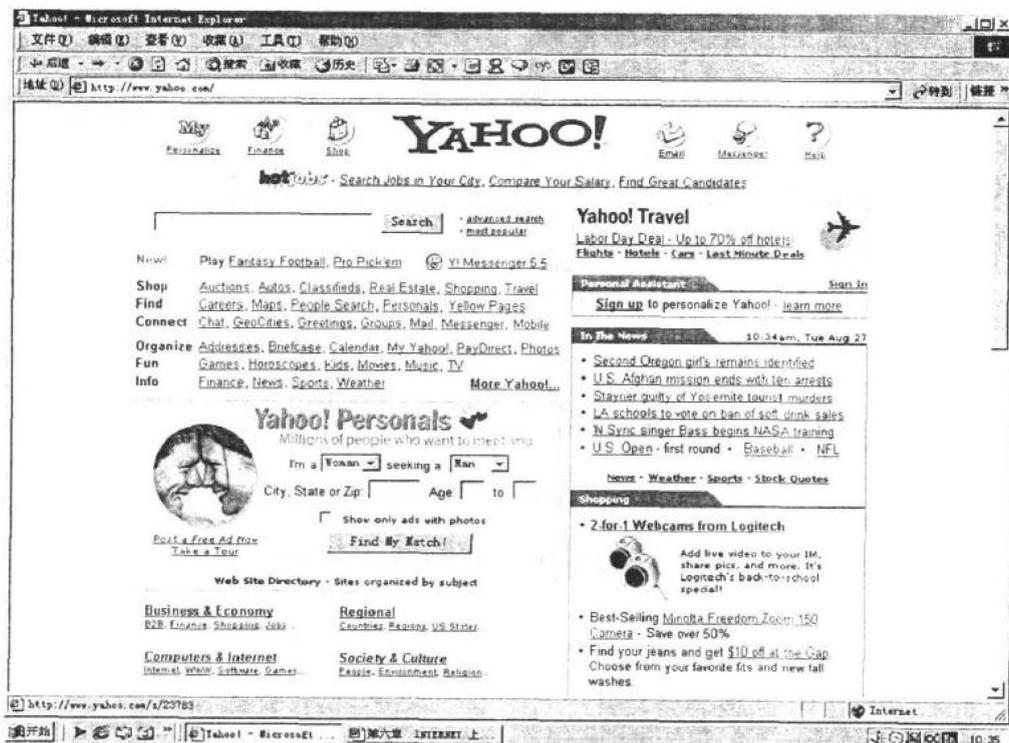


图 1.9 Yahoo 主页