

网络信息资源检索与利用 系列丛书

张厚生 华薇娜 主编



气象水文网络信息资源 检索与利用

田红梅 编著
东南大学出版社

内 容 提 要

本书针对气象水文专业人员的需要,论述了因特网上气象水文信息资源及其组织与检索的基本方法,详细介绍了各种气象水文网络信息资源的存在形式、获取途径、检索方法等,并对现有的网络信息资源进行了分析和评价。全书共分九章,内容包括网络信息资源概述、网络信息资源的组织与检索、气象水文相关网络数据库及其检索、气象水文网络电子期刊及其检索、气象水文网络电子图书及其检索、气象水文网络特种文献资源及其检索、气象水文网站资源及其检索、数字图书馆气象水文信息资源检索以及网络信息资源的分析、评价与利用。

本书可作为高等院校相关专业研究生、本科生教学用书,也可作为气象水文专业人员的参考书、工具书,对于使用因特网进行相关专业信息资源检索的科研人员、爱好者等也具有较高的参考价值。

图书在版编目(CIP)数据

气象水文网络信息资源检索与利用/田红梅编著.

南京:东南大学出版社,2005.9

(网络信息资源检索与利用系列丛书/张厚生,华薇娜主编)

ISBN 7-5641-0151-2

I. 气... II. 田... III. 水文气象学—计算机
网络—情报检索 IV. G354.4

中国版本图书馆 CIP 数据核字(2005)第 080514 号

气象水文网络信息资源检索与利用

出版发行 东南大学出版社
出 版 人 宋增民
社 址 南京市四牌楼 2 号(邮编:210096)
电 话 (025)83792154 (025)83362442(传真)
电子邮件 gujinliang@etang.com
印 刷 溧阳市晨明印刷有限公司
开 本 700mm×1000mm 1/16
印 张 18.5
字 数 383 千字
版 次 2005 年 9 月第 1 版 2005 年 9 月第 1 次印刷
印 数 3000 册
定 价 26.80 元

●东大版图书若有印装质量问题,请直接向发行部调换,电话:025—83795801。

总 序

重视信息素养培训,开展信息检索和利用教育,培养高素质的建设人才,面向在校大学生讲授信息检索课,已成为我国大学用户教育的一大特色。

自从国家教育部文件《关于在高等学校开设“文献检索与利用”课程的通知》发布后,近 20 年来,我国高等学校的文献检索与利用教育得到了较快的发展,相当多的院校已基本上形成了新生入学时的“入馆”教育、本科生的文献检索课程教育、研究生的专业信息资源查找和利用教育的教学体系。进入 21 世纪以来,一些学校适时地开设了网络信息检索与利用的课程,如北京大学开讲了供全校公共选修的“电子资源检索与利用”课,并将该课程分为日常培训课“一小时讲座系列”和专场讲座,其专场讲座的学生覆盖率达到 2/3。

高校信息检索课程的建设与发展,对培养学生的自学能力、获取信息的能力、动手能力和创新能力具有积极的作用。信息检索课程教学具有专业性、实践性等基本特点。所谓专业性,是指各个学校结合本校的系科专业,讲授带有明显专业色彩的信息资源检索与利用的途径、方法;所谓实践性,是指这门课重于实际资讯查找,重于方法和技能的训练,强化课程中的实践检索环节,把理论知识联系到操作应用、落实到课题查找分析的实际过程中。

近 20 年来,信息检索的手段也在飞速地演变。随着社会的进步、科学的发展、信息新技术的迭出,新的知识内容、检索手段不断产生,传统的媒体和检索工具、检索方式也在不断发生变化,人们的信息意识、信息观念在不断更新,经济营造和科学技术的创新能力在不断增强。随着 Internet 和网络技术、多媒体技术的发展,数字化信息资源的涌现,网络环境给我们展现了一个五彩缤纷的信息世界。更大数量级的信息,更多的信息类型和信息表现形式,以及更复杂的时空关联关系,海量的网上信息资源既为信息的开发与利用提供了便利条件,也为信息的发布与分享提供了外部环境;然而,信息产生和流动的随机性、信息时空关系和系统状态的不确定性,导致查找和使用上的困难。数字化、网络化信息的分散、无序、动态变化,以及信息的庞杂同用户特定需求之间的矛盾也给人们检索与利用信息带来了障碍和不便。面对这些状况,为适应信息检索的发展要求,我们特地组织了相关的专家学者,编写了本套网络信息资源检索与利用系列丛书。

《网络信息资源检索与利用系列丛书》由南京大学、东南大学、河海大学、南京师范大学、南京航空航天大学等多个单位的从事信息检索教学的老师和有关专家分头编写,一册以一两个学科、专题为重点,希望避免千篇一律的“大锅菜”式的做法,而专门围绕某一两个学科专业来做文章,深入讲述专业学科的网络信息检索、

分析与使用。已拟订的丛书分册有:生物学、中学、经济学、法学、机电、土木工程、城建环保、医药等。每分册将对相关学科的网络信息的生成、传播、存贮、检索的技术方法作较为详细的阐述。为便于读者在课程教学中的实际操作应用,本丛书侧重于介绍各相关专业的数据库检索,网络信息检索工具的使用,专门的搜索引擎及其使用,信息获取与应用实例等。可以说,这是一套建立在实践基础上的多学科多分册的网络信息检索与利用方面的实用性教材。

建立在思考基础上的著作,产生视角,产生观点;建立在研究基础上的著作,产生知识,产生理论;建立在操作实践基础上的教材,便于学习,便于推广和应用。本丛书可作为本科生、研究生的授课教材,也可为上网的网民用户提供必要的帮助和指导。当前,越来越多的人正在从网上发现或挖掘着自己想要找寻的专业学科信息,正如美国学者 Christine L. Borgman 教授在《从古腾堡到全球信息基础设施——网络世界中信息的获取》(From Gutenberg to the Global Information Infrastructure)一书中所说:“网络群体将促进使所有人都受益的信息社会的发展。通过人际信息交流的改善,人类会有更多的和平、友谊和合作;通过信息的获取,人们可以获得更多的教育、商业和社会利益;通过技术日益先进的工作环境,人们的劳动会有更多的效率;通过在全球化市场的公平竞争,人们的经济会日益发展。”网络信息的检索与利用,给人们带来的社会效益和经济利益是无限的。《网络信息资源检索与利用系列丛书》的编写出版是适时的,符合人们上网的参考需要,当然是值得欢迎的。

网络信息变化很快。要与时俱进、推陈出新,希望各专题分册在每过一定的时段之后,能够及时修订增补,不断以新的内容奉献给莘莘学子,并让读者大众得以分享。

相信读者在阅读这套丛书时会有自己的发现和见解。

张厚生

2004年2月20日于东南大学

前 言

在信息化社会里,信息已成为一种重要的战略资源。而科技资源作为一种重要的信息资源对科学技术的发展至关重要。随着气象水文科学技术的迅速发展,特别是与计算机网络技术的结合,气象水文成果层出不穷,致使气象水文信息资源的数量在急剧增加。气象水文信息是国家基础信息的重要组成部分,近几年来国家投入了大量的人力、物力和财力,建成了气象水文信息数据库、资源网站、网络服务系统等等网上资源,在国民经济建设中发挥了巨大作用。但是,由于长期以来人为形成的数据壁垒,科技领域的重复建设问题令人触目惊心。据《科学时报》报道,北京某著名大学,两个系只是一路之隔,却各建了一个 NOAA 卫星数据接收站;仅长春市 2000 年就已有 4 个 NOAA 接收站,建一个这样的接收站经费在 10 万美元到 100 万美元之间。可见,科技资源不能共享不仅造成了国家投资的巨大浪费,而且在相当程度上制约了我国科研水平的提高。

然而,信息资源共享是建立在共知的基础上的,没有共知就不可能共建,更不可能共享。

互联网上的信息浩如烟海,要快速有效地获取所需要的信息,不仅需要大量的实践,更需要有专门的知识。本书针对气象水文专业人员的需要,论述了因特网上气象水文信息资源及其组织与检索的基本知识与方法,并通过对目前网上已经建立的、具有一定代表性的气象水文信息资源网站、相关数据库、服务系统等专业学科信息的发现和挖掘,介绍了气象水文网络信息资源的存在形式、获取的途径与技术,目的就是让专业及相关人员了解和掌握气象水文信息资源的基本状况、检索技能和利用方法,并为他们提供必要的检索帮助和指导,以提高现有气象水文信息资源的使用效益,使人们在共知信息资源的基础上充分共享信息资源。

本书的特点,一是专业性强,在阐述一般网络信息资源检索的基本理论的基础上,以大量篇幅着重介绍气象水文网络信息资源及其检索的技术与方法;二是内容全面,既介绍了各类网络数据库资源,又介绍了网络特种信息资源,还介绍了各类网站资源、电子期刊、电子图书以及数字图书馆资源;三是可操作性强,读者凭借书中介绍的检索知识和搜索引擎的检索方法,以及所提供的资源站点等,去检索、选择、获取自己学习、研究所需的网络信息,从中肯定会受益匪浅。

由于时间仓促,水平有限,欠妥之处在所难免,真诚地期望得到读者和专家的批评指正!

田红梅

2005 年 7 月

目 录

■ 网络信息资源概述	(1)
1.1 Internet 与网络信息资源	(1)
1.1.1 Internet 的发展	(1)
1.1.2 网络信息资源的概念	(2)
1.1.3 网络信息资源的特点	(2)
1.1.4 网络信息资源的类型	(3)
1.2 我国网络信息资源的分布及其状况	(6)
1.2.1 域名统计	(6)
1.2.2 网站统计	(6)
1.2.3 网页统计	(8)
1.2.4 在线数据库统计	(9)
■ 网络信息资源的组织与检索	(13)
2.1 网络信息资源的组织	(13)
2.1.1 网络信息组织的概念	(13)
2.1.2 网络信息资源的组织方式	(13)
2.1.3 网络信息资源的组织方法	(16)
2.1.4 网络信息组织的技术	(26)
2.1.5 网络信息组织的发展趋势	(28)
2.2 网络信息资源的检索	(31)
2.2.1 网络信息检索的特点	(31)
2.2.2 网络信息检索的方法与技术	(32)
2.2.3 网络信息检索策略	(33)
2.2.4 网络信息检索效果评价	(36)
2.3 网络搜索引擎及其使用	(38)
2.3.1 搜索引擎的概念	(38)
2.3.2 搜索引擎的产生与发展	(39)
2.3.3 搜索引擎的类型	(39)
2.3.4 搜索引擎的工作方式与检索方法	(40)

2.4 常用搜索引擎简介	(42)
2.4.1 百度(Baidu)	(42)
2.4.2 网易(NetEase)	(44)
2.4.3 搜狐(sohu)	(45)
2.4.4 Google	(46)
2.4.5 Yahoo	(49)
2.4.6 Altavista	(52)
2.4.7 Excite	(55)
2.4.8 ProFusion	(56)
2.4.9 Dogpile	(57)
 气象水文相关网络数据库及其检索	(59)
3.1 网络数据库	(59)
3.1.1 网络数据库的基本概念	(59)
3.1.2 网络数据库的特点	(59)
3.1.3 网络数据库的基本构成	(62)
3.1.4 网络数据库的类型	(62)
3.2 主要网络数据库及其检索	(64)
3.2.1 综合性网络数据库	(64)
3.2.2 专业网络数据库	(81)
 气象水文网络电子期刊资源及其检索	(89)
4.1 网络电子期刊与开放存取	(89)
4.1.1 网络电子期刊的概念	(90)
4.1.2 网络电子期刊的开放存取	(90)
4.2 网络电子期刊的类型与特点	(92)
4.2.1 网络电子期刊的类型	(92)
4.2.2 网络电子期刊的特点	(93)
4.2.3 网络电子期刊目前存在的问题	(94)
4.3 国内外主要网络电子期刊资源及其检索	(95)
4.3.1 国内主要电子期刊资源及其检索	(95)
4.3.2 国外主要电子期刊资源及其检索	(102)
4.3.3 免费网络全文电子期刊资源及其检索	(111)
4.3.4 气象水文电子期刊及其站点	(118)
 气象水文网络电子图书及其检索	(125)
5.1 网络电子图书的定义与发展	(125)

5.1.1	网络电子图书的定义	(125)
5.1.2	网络电子图书的发展	(126)
5.2	网络电子图书的特点与类型	(127)
5.2.1	电子图书的特点	(127)
5.2.2	电子图书的类型	(128)
5.3	气象水文网络电子图书的检索	(131)
5.3.1	通过商业公司出版的电子图书数据库检索	(131)
5.3.2	通过图书馆的 WebPAC 检索	(135)
5.3.3	通过专业搜索引擎检索	(138)
5.3.4	通过书业网站检索	(140)
	气象水文网络特种文献资源及其检索	(143)
6.1	会议网络信息资源检索	(143)
6.1.1	会议文献概述	(143)
6.1.2	会议文献数据库检索	(144)
6.1.3	会议信息资源网站及其检索	(149)
6.2	科技报告网络信息资源检索	(152)
6.2.1	气象水文科技报告概述	(152)
6.2.2	科技报告数据库检索	(155)
6.2.3	科技报告资源网站及其检索	(157)
6.3	专利网络信息资源检索	(158)
6.3.1	气象水文专利文献概述	(158)
6.3.2	专利文献数据库检索	(159)
6.3.3	专利信息资源网站及其检索	(164)
6.4	学位论文网络信息资源检索	(166)
6.4.1	学位论文概述	(166)
6.4.2	学位论文数据库检索	(166)
6.4.3	学位论文网站资源及其检索	(171)
	气象水文网站资源及其检索	(174)
7.1	政府机构网站	(174)
7.1.1	中国气象局	(174)
7.1.2	国家海洋局	(176)
7.1.3	中华人民共和国水利部	(177)
7.1.4	美国国家海洋与大气管理局	(177)
7.1.5	加拿大气象局	(178)
7.1.6	日本气象厅	(179)

7.2 国际组织机构网站	(179)
7.2.1 世界气象组织	(179)
7.2.2 国际气象学与大气科学协会	(181)
7.2.3 国际海洋物理科学协会	(183)
7.2.4 海洋研究科学委员会	(183)
7.2.5 国际水文科学协会	(183)
7.3 学会、协会机构网站	(183)
7.3.1 国内气象水文学会、协会网站	(184)
7.3.2 国外气象水文学会、协会网站	(185)
7.4 气象水文教育机构网站	(190)
7.4.1 国内教育机构网站	(190)
7.4.2 国外气象水文教育网站	(194)
7.5 气象水文科研机构网站	(196)
7.5.1 国内气象水文科研机构网站	(197)
7.5.2 国外气象水文科研机构网站	(201)
7.6 气象水文专题网站与门户网站	(203)
7.6.1 专题网站	(203)
7.6.2 门户网站	(206)
7.7 气象水文信息服务网站	(209)
7.7.1 国内气象水文信息服务网站	(209)
7.7.2 国外气象水文信息服务网站	(214)
 8 数字图书馆气象水文信息资源检索	(216)
8.1 数字图书馆的定义、特点与功能	(216)
8.1.1 数字图书馆的定义	(216)
8.1.2 数字图书馆的特点	(218)
8.1.3 数字图书馆的功能	(219)
8.2 数字图书馆的发展与服务	(219)
8.2.1 数字图书馆构想的提出	(219)
8.2.2 数字图书馆发展的几个阶段	(220)
8.2.3 数字图书馆的启动与实施	(220)
8.2.4 数字图书馆的信息服务模式与实现技术	(222)
8.3 数字图书馆气象水文信息资源检索	(225)
8.3.1 数字图书馆信息资源及其服务	(225)
8.3.2 数字图书馆的气象水文资源检索	(226)
8.3.3 网上图书馆的气象水文资源检索	(230)

9 网络信息资源的分析、评价与利用	(238)
9.1 网络信息分析	(238)
9.1.1 网络信息分析的含义	(238)
9.1.2 信息分析的特点与类型	(239)
9.1.3 信息分析的功能与作用	(240)
9.1.4 信息分析的方法	(240)
9.1.5 信息分析的流程	(246)
9.2 网络信息的采集	(247)
9.2.1 网络信息采集的原则	(247)
9.2.2 网络信息采集的方法	(248)
9.2.3 网络信息采集的步骤	(251)
9.3 网络信息的鉴别	(252)
9.3.1 信息的整理	(253)
9.3.2 信息的鉴别	(254)
9.4 网络信息资源的评价	(255)
9.4.1 网络信息资源评价的指标	(255)
9.4.2 网络信息资源评价的方法	(256)
9.4.3 网络信息资源评价应注意的问题	(258)
9.5 利用网络信息资源进行论文写作	(258)
9.5.1 研究报告的撰写	(259)
9.5.2 学术论文的撰写与投稿	(261)
9.5.3 学位论文的撰写	(267)
9.6 网络信息资源与知识产权保护	(272)
9.6.1 知识产权的概念	(272)
9.6.2 知识产权保护的意义	(273)
9.6.3 网络信息资源知识产权保护的内容	(274)
9.6.4 知识产权保护技术	(276)
参考文献	(279)
参考网站	(280)

1.1 Internet 与网络信息资源

1.1.1 Internet 的发展

信息技术革命强烈地改变着世界,计算机产业的蓬勃发展给各国带来了巨大的机遇和挑战,尤其是网络技术的广泛应用,模糊了国与国之间的边界,缩短了人们之间的距离。正是在这种情况下,以美国为首的发达国家纷纷提出了各自建设信息高速公路的国家政策,力图能够把握今后社会发展的方向。亚洲的主要国家也正以巨大的热情响应这一决策,纷纷提出建设国家信息基础设施的计划,以适应信息社会发展的需要。Internet 是信息高速公路的基础和入口点,它正潜移默化地改变着人类的生活方式。

就全世界而言,没有人能知道 Internet 目前的确切规模。因为除了运行 TCP/IP 通信协议的网络外,还有一些并非基于 IP 通信协议的网络(如 BITNET 和 DECnet 等)为方便其用户与 Internet 用户交换电子邮件,也通过网关(Gateway)与 Internet 连通。

与发达国家相比,中国的网络发展起步较晚,不过发展速度还是比较快的。特别是进入 20 世纪 90 年代后,中国的互联网获得了突破性的发展,1990 年 11 月,中国注册了自己的顶级域名 CN;1992 年 6 月,在日本的神户,中美首次讨论中国连入 Internet 问题,同年 12 月底,作为全国一流大学的清华大学建成了中国第一个基于 TCP/IP 体系结构的校园网络——清华大学校园网;1994 年 4 月中国连入 Internet 的 64 K 国际专线开通,成为第 77 个连入 Internet 的国家;1994 年 5 月,中科院计算机网络信息中心的网络域名服务器投入使用,结束了中国顶级域名一直在国外的历史。

根据中国互联网络信息中心(CNNIC)2005 年 1 月 19 日发布的“第十五次中国互联网络发展状况统计报告”,我国上网用户总数为 9 400 万,比半年前增长 8%,其中使用宽带上网的人数达到 4 280 万;上网计算机达到 4 160 万台,比半年前增长了 14.6%;CN 下注册的域名数、网站数分别达到 43 万和 66.9 万,分别比半年前增长了 5 万和 4.2 万;网络国际出口带宽总数达到 74 429 M,IPv4 地址总数达 59 945 728 个,分别比半年前增长 38%和 21%。

随着世界各国信息高速公路计划的实施,Internet 主干网的通信速度还将大幅度提高。有线、无线等多种通信方式将更加广泛、有效地融为一体,Internet 的

商业化应用将大量增加,商业应用的范围也将不断扩大,Internet 的覆盖范围、用户入网数将以令人难以置信的速度发展。Internet 的管理与技术将进一步规范化,使用规范和相应的法律规范正逐步健全和完善。

然而,信息是网络的灵魂,没有信息,网络就没有任何价值。如果 Internet 上没有丰富的信息,则 Internet 再大也是一个空网。信息是 Internet 的重要组成部分,信息资源的共享是建立 Internet 的最主要目的之一,信息资源的丰富和通信的快捷方便是 Internet 迅速发展的动力。总之,人类社会必将更加依赖 Internet,人们的生活方式将因此而发生根本的改变。

1.1.2 网络信息资源的概念

目前,网络信息资源还没有一个统一的定义,关于其定义的表述多种多样。有人认为,网络信息资源是指从计算机、通信技术、多媒体技术相互融合而形成的网络上可以查找到的资源;也有人认为网络信息资源是指以电子数据的形式将文字、图像、声音、动画通过计算机网络再现出来的可以利用的各种信息资源的总和;还有人认为网络信息资源是以数字化形式存储,并借助计算机与网络通讯设备收集、生产、传递的信息资源;中国互联网络信息中心将网络信息资源定义为互联网络上公开发布的网页和在线数据库的总和,等等。我们认为,网络信息资源是电子信息资源的一种类型,它以电子数据的形式将各种信息(文字、图像、声音、动画等)存储在光、电、磁等载体上,通过网络设备、计算机或终端等方式传递、再现。

在网络信息系统中,信息是核心资源,是网络传播的主要形式,是网络用户上网搜索的主要对象。人力资源、技术资源、资金资源、资本资源为支撑性资源,由于网络信息资源的开发与利用以其为基础和条件,因此它们的地位更加突出。

1.1.3 网络信息资源的特点

网络信息资源的生产、传播通过 Internet 来实现,具有以下特点:

1) 以网络为传播媒体

传统意义上的信息源是指存储信息的载体,如纸张、报刊杂志等,随着信息技术的进步,进而出现了磁盘、磁带、光盘等。在网络时代,信息的存储要借助于一种不同于以往载体形式的新型载体——网络,与传统形式上的信息源——一种实实在在的东西不同,读者面对的是计算机屏幕上所显示的来自 Internet 上各种服务器的各种信息。

2) 以多媒体为内容特征

除文本信息外,网络上还包含大量非文本信息,如图像、声音、动画等多媒体信息。Internet 上的电子信息资源,其处理采用多媒体技术,大多是图、文、声并茂,集报刊、广播、电视的优势于一体。这要归结于计算机硬件的发展、图像处理与压缩技术、网络传输速率的提高等诸多方面。

3) 传播方式具有多样性、交互性

因特网具有传播信息形态多样、迅速方便、全球覆盖、自由交互的特点。报刊、广播、电视三大传媒所具有的一切表现形式和特点互联网可以兼备,互联网更具备三大传媒所不具备的一些特点,成为知识经济时代最具发展潜力的大众传媒。互联网打破了时空的限制,打破了传统的地缘政治、地缘经济、地缘文化的概念,形成了以传播信息为中心的跨国界、跨文化、跨语言的全新传播方式,不断地更新着人们的观念,丰富着人们获取信息的手段,影响着人们的生活。

4) 数量巨大,增长迅速

20 世纪 90 年代以来,由于因特网的商业化、WWW 和浏览器以及各种搜索引擎的出现,网络信息数量迅速增长,网上实时信息增多,大量的时效性信息潮水般涌来。在网上出版发布信息的信源广泛,涉及各行各业、各领域的机构、团体、企业、单位乃至个人。由于网络带宽的增加,网络信息传递速度也越来越快,时效性越来越强。据统计因特网每天发布 14 万件新信息,全网信息总量达 20TB。

5) 信息共享程度高

由于信息存储形式及数据结构具有通用性、开放性和标准化的特点,在网络环境下,时间和空间范围得到了最大程度的延伸和扩展。一份信息资源上网后,不仅可以及时地提供给本地用户,而且可以发散到全球各个角落,用户不需排队等候就可以共享同一份信息资源。

6) 使用成本低

网络信息资源绝大部分可免费使用,用户所需支付的主要是市话费和网络费用。低费用的网络信息资源有效地刺激了用户的信息需要,从需求的角度拉动了网络信息资源的有效、合理配置。

7) 动态性强,缺乏管理

由于受时效与各种因素的影响,网络信息内容、所在地址、链接关系等都处于动态变化之中,生产、更替、消亡情况难测。另外,信息的发布有很大的自由度和随意性,信息污染严重,如失实信息、虚假信息、重复信息、过时信息、缺损信息、误差信息、噪声信息等,与有用信息、高质量信息鱼龙混杂,缺乏必要的过滤、质量控制和管理控制,信息资源良莠不齐。

1.1.4 网络信息资源的类型

对网络信息资源的类型,人们可以从不同角度进行不同的划分。美国联机计算机图书馆(OCLC)的“因特网资源计划”根据文件类型将网络信息资源划分为:政府出版物、百科全书、电子论坛、书目、索引和摘要、图书馆目录、字词典、公告、导览、宣传资料等几十种。国际数据集团(IDG)1996 年出版的《万维网指南》依据学科领域将网络信息资源划分为:农业、人类学、天文学和商业等 57 个类目。还有人从出版发布方式将网络信息资源分为:非正式出版的网络信息、正式出版的电子出

出版物、半正式出版的网络信息,等等。

1) 按信息存储或发布形式划分

(1) 联机数据库 各类数据库是联机存取电子信息源的主体。据统计,仅至1991年,世界上就有数据库7 600个。目前,很多重要数据库都已连入因特网,仅DIALOG系统就有600多个联机数据库。这些数据库涉及各个学科、各种文献。

(2) 联机馆藏目录库 网上已有6 000多个电子图书馆,包括美国国会图书馆在内的600多所著名公共图书馆、大学图书馆及4 000多个学术机构的馆藏机读目录库,通过网络对外开放,提供OPAC(联机公共检索目录)服务。

(3) 电子图书 在因特网上电子图书越来越多,电子图书涉及的领域非常广泛,包括文学、艺术、科学、人文等各个领域。其中,网上电子词典、百科全书更是独树一帜。

(4) 电子期刊 在因特网上,有数万种电子期刊向用户提供服务。世界著名的*Times*、*Forbes*、*Science*、*Nature*等都有网络电子版。《神州学人》是我国上网最早的网络电子版期刊之一。

(5) 电子报纸 据统计,网上已有上千种报纸供用户使用,其中很多报纸免费向用户提供。如美国的《华尔街日报》、《华盛顿邮报》、《纽约时报》及我国的《人民日报》、《光明日报》、《经济日报》等报纸都已在网上发行。

(6) 软件与娱乐游戏类。

(7) 网上教育培训信息、旅游景点介绍、交通信息、天气预报、股票行情、BBS、广告、网络新闻等动态信息等。

(8) 电子邮件、专题讨论组、论坛、电子会议、电子布告新闻、文件报告、统计数据、法规条例等等。

2) 按信息的表现形式划分

(1) 文本信息资源 一般的文本信息资源是按知识单元的线性顺序排列的,在检索时存在较大的局限性和片面性。

(2) 超文本信息资源 超文本信息资源是按知识单元及关系建立的知识结构网络,它通过网络上各节点的链路把相关信息有机地结合在一起。检索超文本信息资源时可从任何一个节点开始,以知识片段及其关系作为检索、追踪的依据。

(3) 多媒体信息资源 是包括文本、图像和声音等多种信息表达方式的总称。

(4) 超媒体信息资源 超媒体信息资源具有超文本和多媒体两种信息资源的特点。在超媒体信息系统中,不同类型的媒体信息能高度综合和集成,空间上图、文、声并茂,时间上多媒体信息同步实现。

3) 按所用网络传输协议划分

(1) 万维网(World Wide Web,简称WWW或Web)信息资源 它是建立在超文本传输协议(HTTP)和超文本、超媒体技术基础上,集文本、图像、声音、动画等多媒体信息为一体,以超文本标记语言描述、存储、检索,以直观网页用户界面展

现、提供信息的网络信息资源。它使用简单方便,功能强大,发展迅速,成为因特网信息资源的主流。

(2) 远程登录(Telnet)信息资源 它是借助于远程登录(Remote Login),在网络通讯协议 Telnet(Telecommunication Network Protocol)的支持下,在远程计算机上登录,使自己暂时成为远程计算机的终端,从而可以实时访问、共享、使用远程计算机中对外开放的资源。这些资源既包括硬件资源,也包括软件、大型数据库等信息资源。

(3) FTP 服务器信息资源 FTP(File Transfer Protocol)是因特网使用的文件传输协议,它的主要功能是完成因特网的联网计算机之间的文件传输。任何以计算机方式存储的信息均可以保存在 FTP 服务器中,因此通过 FTP 服务器可获取的网络信息资源很广泛。Archie 是帮助查找在 FTP 服务器上存放文件的检索系统,用户可用关键文件名查询,以确定文件的位置,目的是检索存储在网络上可供检索的公共文档中的软件,发现哪些文件可通过 FTP 服务器向公众提供。

(4) Gopher 信息资源 Gopher 是一种基于菜单的网络信息服务资源 Gopher 服务器,存储了各种各样的信息,信息被列入菜单等生成过程中要经过一定的审核,每个 Gopher 服务器内部都存放本部门或本地区用户感兴趣的信息。Gopher 服务器还可以提供如 WWW、FTP、Telnet、Archie、WAIS 等的连接。其实质是一个按照等级式菜单将不同主题的文本信息组织起来的查询系统,用户可以依据主题指南逐层浏览和查询信息。它以主题目录的方式对网络中的文本信息进行组织,使用户根据按等级排列的菜单系统巡视网络资源。该系统由于以菜单驱动界面,使用起来非常简单。

(5) 网络用户服务组信息资源 网络用户通信或服务组包括:新闻组(Usenet Newsgroup)、邮件列表(Mailing List)、专题讨论组(Discussion Group)、兴趣组(Interest Group)等,实质上是一组对某一特定主题有共同兴趣的用户组成的电子论坛,在论坛中交流的文章是一封封 E-mail,人们通过它能便捷地进行多次信息交流。它主要包括某一专题领域的新闻、研究动向、最新成果发布、交谈、解释疑惑、讨论、评论等,其信息交流非常广泛而直接。

(6) WAIS 信息资源 广域信息服务系统(WAIS)是一种易于检索并可获取远程数据文档的动态超文本系统,通过利用网络信息检索协议(network information retrieval protocol)和单一的用户友好界面检索网络上分布在不同地方的信息,特别适于检索文本文件。使用 WAIS,用户有可能阅读世界各国的报纸,扫描各种专业数据库。

4) 按信息的内容划分

可分为军事信息、政务信息、经济信息、科技信息、管理信息、文体信息等不同领域的网络信息。

1.2 我国网络信息资源的分布及其状况

基于目前我国域名、网站、网页和网络数据库等信息资源快速增长,2005年4月14日,中国互联网络信息中心发布了《2004年中国互联网络信息资源数量调查报告》。报告显示,2004年我国网络信息资源在数量、质量和服务方面都取得了新的进展,保持健康发展的态势。这些数据反映了我国目前网络信息资源的分布状况,同时也为我国网络信息资源的建设提供了良好的依据。

据统计,截止到2004年12月31日,全国域名数约为185.23万个,比上年增长56.0%,其中广东、北京居前两位;全国网站数约为66.89万个,比上年增长12.3%,其中COM和COM.CN比例最高;网页总字节数增长最快,同期相比增幅为238%。网页总数为6.5亿个,同期相比增长108.6%。平均每个网站的网页数为1297个,同期相比增长147.5%。全国在线数据库约为30.6万个,比上年增长80.1%。

1.2.1 域名统计

截止到2004年12月31日,全国域名数为1852300个。

(1) 所有域名中,以COM.CN和COM结尾的最多,占域名总数的71.6%,其次分别为以NET.CN和NET结尾的域名占12.5%,直接在顶级域名CN下注册的二级域名占10.5%。

(2) 在地域分布中,广东(16.3%)、北京(12.9%)、福建(11.2%)位列全国前三名。

1.2.2 网站统计

截止到2004年12月31日,全国网站总数约为668900个。

(1) 网站WWW性质分布,其中COM和COM.CN比例最高占71.3%。具体见表1-1。

表1-1 网站WWW性质分布

	COM (COM .CN)	NET (NET .CN)	ORG (ORG .CN)	GOV .CN	EDU .CN	AC .CN	行政区域. CN	.CN	总计
比例	71.3%	11.9%	3.3%	1.5%	略	0.1%	0.4%	11.5%	100.0%

(2) 网站数排在前5位的省市依次是北京、广东、浙江、上海和江苏,这5个省市的网站总数占全国总数的64.6%。如表1-2是网站地域分布。

表 1-2 2004 年网站地域分布

北京	天津	上海	重庆	河北	山西	内蒙古
18.7%	1.0%	8.7%	1.2%	2.5%	0.7%	0.4%
辽宁	吉林	黑龙江	江苏	浙江	安徽	福建
3.3%	0.7%	0.9%	7.5%	11.5%	1.7%	5.7%
江西	山东	河南	湖北	湖南	广东	广西
1.1%	4.0%	2.0%	2.3%	1.2%	18.2%	1.2%
海南	四川	贵州	云南	西藏	陕西	甘肃
0.4%	1.9%	0.4%	0.7%	0.3%	0.8%	0.4%
青海	宁夏	新疆				
0.1%	0.2%	0.3%				

(3) 各类网站中,企业网站数的比例最大,占网站总数的 60.7%,其次为个人网站,占 13.5%,第三是商业网站,占 11.4%,随后依次为其他公益性网站占 4.6%,教育科研网站占 4.6%,政府网站占 3.6%,其他网站占 1.6%。

(4) 关于各类网站每天页面访问数(页面个数),见 1-3 表。

表 1-3 各类网站每天页面访问情况

% Pageview	政府 网站	企业 网站	商业 网站	教育科研 网站	个人 网站	其他公益性 网站	其他 网站	总体
50 以下	33.3	45.8	33.7	26.8	45.5	22.7	11.1	40.3
51~200	30.3	26.9	17.9	31.7	23.5	27.3	44.4	25.7
201~1 000	21.2	17.1	25.3	24.5	19.6	18.2	33.4	19.7
1 001~5 000	15.2	7.1	14.7	14.6	6.1	22.7	11.1	9.9
5 000 以上	—	3.1	8.4	2.4	5.3	9.1	—	4.4

(5) 关于网站链接情况。链接数是指该网站在自己网站上提供与其他网站的链接数量,报告显示有近半数(48.3%)的网站上没有提供与其他网站的链接。见图 1-1。

(6) 网站信息最主要的来源是自己采集,比例高达 86.9%;21.1%的网站从其他网站处获取信息;有 25.6%的网站信息来源于传统媒体和商业企业提供。

(7) 关于网站提供的语种/文字阅读情况,调查显示,96.3%的网站提供简体中文内容,28.3%的网站提供英文内容。

(8) 全部网站中,有 31.8%的网站提供全站信息搜索服务,95.7%的网站上有联系方式。