

普通高等院校教材

网络信息检索

吕建新 主编
陈兰杰
李哲

陈琳 副主编
侯鹏娟

煤炭工业出版社

普通高等院校教材

网络信息检索

主 编 吕建新 陈兰杰 李 哲
副主编 陈 琳 侯鹏娟

煤炭工业出版社

·北 京·

图书在版编目 (CIP) 数据

网络信息检索/吕建新, 陈兰杰, 李哲主编. --北京: 煤炭工业出版社, 2011

普通高等院校教材

ISBN 978-7-5020-3905-9

I. ①网… II. ①吕… ②陈… ③李… III. ①网络检索 - 高等学校 - 教材 IV. ①G354.4

中国版本图书馆 CIP 数据核字 (2011) 第 148576 号

煤炭工业出版社 出版
(北京市朝阳区芍药居 35 号 100029)

网址: www.cciph.com.cn

煤炭工业出版社印刷厂 印刷
新华书店北京发行所 发行

开本 787mm × 1092mm¹/₁₆ 印张 12¹/₄
字数 284 千字

2011 年 8 月第 1 版 2011 年 8 月第 1 次印刷
社内编号 6715 定价 30.00 元

版权所有 违者必究

本书如有缺页、倒页、脱页等质量问题, 本社负责调换

内 容 提 要

本书立足于网络信息资源的检索与利用这一主题，介绍了网络信息资源检索的基本理论和基本方法，以及常用的网络信息检索系统。

本书可作为高校信息管理与信息系统、电子商务、工商管理、图书馆学、档案学、管理学、编辑出版发行学等专业的教材，也可作为公共选修课的教材，还可供广大信息工作者和科研人员参考。

前 言

网络信息资源是信息资源的重要组成部分，进行网络信息搜索已成为日常生活、学习和工作的一个重要组成部分。因此，掌握网络信息检索方法与技巧成为信息时代的一项基本技能。本书立足于网络信息资源的检索与利用，介绍了网络信息资源检索的基本理论和基本方法，以及常用的网络信息检索系统。

全书共8章，前2章是网络信息检索的基础理论，后6章结合具体的检索对象介绍了各类网络信息资源的检索。主要内容包括绪论、网络信息检索基础、网络信息检索工具——搜索引擎、电子图书与书目信息检索、网络学科资源导航、开放存取信息资源检索、参考信息的网上检索，以及特种文献信息检索。

全书由吕建新、陈兰杰、李哲起草大纲，由多所学校的教师共同编写而成。各章的编写分工如下：第一章由肖岩（石家庄工程职业学院图书馆）编写；第二章由陈兰杰（河北大学管理学院）编写；第三章由孙玲玲（河北民族师范学院图书馆）编写；第四章由吕建新（河北北方学院图书馆）编写；第五章由侯鹏娟（中国地质大学长城学院图书馆）编写；第六章由李哲（外交学院图书馆）编写；第七章由周文超（包头师范学院）编写；第八章由陈琳（河北省委党校图书馆）编写。全书最后由吕建新、陈兰杰、李哲统稿校对。

由于网络信息资源瞬息万变，网络信息检索系统层出不穷，加之时间仓促且水平有限，对本书可能出现的缺点乃至错误，敬请读者批评指正。另对本书所参考的所有资料的作者表示感谢。

编 者

2011年6月

目 次

第一章 绪论	1
第一节 网络信息检索的一般知识.....	1
第二节 网络信息的产生与发展.....	6
第三节 网络信息的类型与特征.....	8
第四节 网络信息的分布	10
本章小结	14
复习思考题	14
第二章 网络信息检索基础	15
第一节 信息检索与网络信息资源检索的含义	15
第二节 因特网信息资源检索语言	20
第三节 网络信息检索的途径	26
第四节 互联网信息资源检索方法	27
第五节 网络信息资源检索技术	31
第六节 网络信息资源检索的步骤	34
第七节 网络信息资源检索效果的评价及影响因素	38
本章小结	39
复习思考题	39
第三章 网络信息检索工具——搜索引擎	40
第一节 搜索引擎概述	40
第二节 常用搜索引擎推介	43
第三节 搜索引擎检索技巧	63
本章小结	65
复习思考题	65
第四章 电子图书与书目信息检索	66
第一节 电子图书检索	66
第二节 常用电子图书数据库简介	68
第三节 书目信息的网上分布及查询技巧	80
第四节 网上书店	82
本章小结	86
复习思考题	86
第五章 网络学科资源导航	87
第一节 国内重要网络学科资源导航门户	87
第二节 国外重要网络学科导航	95

本章小结·····	102
复习思考题·····	102
第六章 开放存取信息资源检索·····	103
第一节 开放存取的概述·····	103
第二节 开放存取的发展历程·····	107
第三节 国内外重要开放存取资源·····	108
本章小结·····	112
复习思考题·····	112
第七章 参考信息的网上检索·····	113
第一节 参考信息概述·····	113
第二节 百科知识的网上检索·····	114
第三节 人物信息的网上检索·····	119
第四节 地理信息的网上检索·····	124
第五节 机构信息的网上检索·····	130
第六节 统计信息的网上检索·····	137
本章小结·····	140
复习思考题·····	140
第八章 特种文献信息检索·····	141
第一节 标准文献检索·····	141
第二节 会议文献检索·····	149
第三节 科技报告检索·····	156
第四节 学位论文检索·····	162
第五节 专利文献信息检索·····	170
本章小结·····	184
复习思考题·····	184
参考文献·····	185

第一章 绪 论

第一节 网络信息检索的一般知识

一、网络

从概念上讲,网络有狭义和广义的解释,狭义的网络仅指计算机网络,广义的网络是指互联网。

1. 计算机网络

即狭义的网络概念,指理论上常说的计算机网络,是由具有独立功能的多个计算机系统,通过通信线路和设备互联而形成的网络系统。这些计算机通过运行网络软件,可以在网络中的各个计算机之间进行数据通信。计算机网络的功能包括:

(1) 数据传递:在计算机之间快速且可靠地传输数据。

(2) 共享资源:用户不受地理位置及时间的限制,可以随时随地使用网络中的数据、软件和硬件设备。

(3) 提高系统的可靠性:一旦系统中的某台计算机发生故障,预先设置的备用计算机可以迅速接替它的工作。

(4) 负载平衡:当某个计算机工作繁重时,可以通过网络把一部分任务分配给相对空闲的计算机去处理,提高了系统的处理效率。

(5) 分布式处理:一项复杂的任务可以分成许多部分,由网络内的计算机分别完成有关部分,使整个系统的功能大为增强。

2. 互联网 (Internet)

即广义的网络概念,也称因特网,是世界上最大的计算机互联通信网络,最早起源于美国国防部的计算机网络 ARPAnet。它本身不是一种具体的物理网络,而是一种虚拟的计算机网络。互联网实际上是把全世界各个地方已有的各种网络,如计算机网络、数据通信网以及公用电话交换网等通过 TCP/IP 协议相互联结,组成一个跨越国界的庞大的综合网络。

按距离远近、地理覆盖范围分类,互联网又包括:

(1) 广域网 (Wide Area Network, 简称 WAN):分布距离大于 50 km,可以覆盖多个单位或多个国家,如我国的 ChinaNET、CERNET 等。

(2) 局域网 (Local Area Network, 简称 LAN):分布距离在 10 km 范围内,通常为某个单位专用,如河北大学校园网等。局域网具有结构简单、投资少、数据传输速度快、可靠性好、保密性强等特点。

(3) 城域网 (Metropolitan Area Network, 简称 MAN):也称都市网,分布距离在 5 ~

50 km 之内，是介于广域网和局域网之间的一种大范围的高速网络，可以专用或公用。

二、网络体系结构与通信协议

在计算机网络中，为了使计算机或终端之间能够正确地传送信息，必须就信息传输顺序、信息格式和信息内容等进行约定。这一整套约定称为网络协议。为了降低协议设计的复杂性，大多数网络按层的方式来组织。层和协议的集合统称为网络体系结构。在计算机网络最初发展的时代，不同的网络，其网络体系结构也是不同的。

为促进网络技术的进一步发展应用，国际标准化组织 ISO 于 1981 年颁布了开放系统互联（Open System Interconnection，简称 OSI）参考模型，也就是人们常说的“七层模型”。它规定了不同计算机网络联网的框架结构，将整个网络的通信功能分为七个层次，即低三层——物理层、数据链路层和网络层，和高四层——传输层、会话层、表示层和应用层。每层间有相应的通信协议，称作同层协议，相邻层之间的通信约束称作接口。OSI 参考模型最终被开发成全球性的网络结构，促使所有的计算机网络走向标准化，具备了互联的条件。

目前在计算机网络中最为通用和广泛接受的协议是 TCP/IP 协议和 IEEE802 协议。

1. TCP/IP 协议

TCP/IP 协议的全称是 Transmission Control Protocol/Internet Protocol，即传输控制协议和网际协议，最初由美国国防部开发并应用于 ARPAnet。其中 TCP 协议对应于 OSI 参考模型的传输层，提供面向连接的可靠传输数据服务，以确保所有传送到某个系统的数据正确无误地到达该系统。IP 协议对应于 OSI 参考模型的网络层，制定了所有在网络上流通的数据包标准，提供了跨越多个网络的单一数据包传送服务。TCP/IP 协议是世界上应用最广的异种网互联的标准协议，是因特网的核心协议。

2. IEEE802 协议

IEEE802 协议由电气电子工程师协会 IEEE（The Institute of Electrical and Electronic Engineer）制定，对应于 OSI 参考模型中的物理层和数据链路层，主要应用于局域网技术。局域网由于低层采用了 IEEE802 协议，可以很容易实现互联。

三、主机与终端

在计算机网络中，主机和终端是进行全网的数据处理和向用户提供网络资源及网络服务的主体设备。

1. 主机（Host）

主机即主计算机系统，负责数据处理和网络控制，为本地用户访问网络或其他主计算机设备，共享资源，以及为网络中其他用户共享本地资源提供服务。主机可以是大型机、中型机、小型机、工作站或微机，能够与其他模块或网络中的主机相连。

2. 终端（Terminal）

终端是网络中用量大、分布广的设备，直接面对用户，是用户进行网络操作、实现人机对话的工具。终端包括简易终端（只具备键盘打字功能）、复合终端（由多台输入输出设备组成）和智能终端（通常为微型机）。

四、客户端/服务器模式

客户端/服务器 (Client/Server, 简称 C/S) 模式, 是互联网的服务模式。在网络出现的早期, 一些公司使用网络技术将远程终端连接到中央计算机上, 这些终端是在中央计算机完全控制下工作的, 这就是所谓的“主从式”(或主仆式) 模式。随着计算机技术的发展, C/S 模式很快发展起来。这种模式就是将网络中需要处理的任务分配给客户端和服务端共同完成, 其基础就是分布式计算, 其中服务器是提供服务的, 客户端是进行访问服务的。

1. 客户端

客户端也可以称为客户机, 主要用于运行客户软件, 与服务器进行交互, 生成一个请求, 通过网络发送请求到服务器, 然后等待回答。客户端通常由个人计算机担当。客户端软件有通用型的, 如 Netscape、Internet Explorer、E-mail 软件等, 适用于任何资源, 也有人称此为浏览器/服务器 (Browser/Server) 模式, 即 B/S 模式; 也有专用的, 只适用于某种类型的资源, 如 Adobe Acrobat Reader, 用于阅读 PDF 文件。

2. 服务器

服务器由一些更为复杂的软件组成, 在接收到客户端发来的请求之后, 分析其请求, 并予以回答, 回答的信息 (数据包) 通过网络发到客户端, 客户端接收到请求后, 再将结果显示给用户。与客户端不同, 服务器的程序必须一直运行, 随时准备接收请求, 客户端可以在任何时候访问服务器。服务器由于负担较重, 一般运行在大型、中型、小型计算机或高配置的微机上, 以便响应多个客户请求, 运行多种服务程序。

五、地址与域名

为了使用户能够在互联网上方便而快捷地找到需要与其连接的主机, 必须解决如何识别网上主机的问题。互联网通过 IP 协议解决了这个问题, 即给每台主机统一分配一个地址, 并隐藏了原来的物理地址, 这个地址即被称为“IP 地址”。

IP 地址由 4 节数字组成, 每节数字的取值范围为 0 到 255, 每节数字之间用“.” 隔开。根据网络规模的大小, IP 地址可分为 A、B、C、D、E 五类, 其中 A、B、C 为主要类型地址。A 类为大型网络, IP 地址第一节作为网络地址, 后三节作为网内主机地址, 可容纳 1600 多万台主机; B 类为中型网络, IP 地址中网络地址和主机地址各占两节, 可容纳 6 万多台主机; C 类为小型网络, IP 地址的前三节用作网络地址, 最后一节为主机地址, 可容纳 254 台主机。

有多少个 IP 地址, 就意味着有多少台计算机在互联网上。通常互联网的服务机构从互联网网络信息中心 InterNIC 那里先注册, 成批得到 IP 地址后, 再依次分配给自己的用户使用。用户使用的 IP 地址可以是固定不变的, 也可以是动态的, 即随机得到某一个 IP 地址段内的某一个主机地址。

有了 IP 地址, 仍存在另一个问题, 就是普通用户记住抽象数字的 IP 地址是很困难的。为了向一般用户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, 也就是域名系统 DNS (Domain Name System), 其结构如下:

主机名 . 机构名 . 网络名 . 顶级域名

例如，清华大学的域名为 `www. tingham. edu. cn`，其结构如下：

WWW 服务器 . 机构名 . 教育机构 . 中国

域名系统是分层管理的，其中顶级域名和第二级域名由 InterNIC 管理。顶级域名包括三类：一是国家，如 `cn` 代表中国，`uk` 代表英国，`jp` 代表日本；二是国际组织，国际联盟、国际组织可在 “. `int`” 下注册，如世界卫生组织（World Health Organization）的域名为 `www. who. int`；三是通用型第一级域名，如 “. `edu`” 表示教育机构，“. `com`” 表示商业公司，“. `info`” 表示信息服务单位，“. `net`” 表示网络服务机构，“. `org`” 表示非营利组织，“. `gov`” 表示政府机构等。中国所有的域名都是在 “. `cn`” 下的二级域名，域名注册由中国互联网信息中心（CNNIC）管理。

域名系统的使用是在域名服务器的支持下完成的，即由域名服务器保存域名信息，负责从域名到主机 IP 地址的解析。例如，某用户输入北京大学图书馆的域名 `www. lib. pku. edu. cn`，此域名即由域名服务器解析成 IP 地址 `162. 105. 138. 207`，找到该机构的主页。

六、其他基本概念

1. 万维网（WWW）

万维网（World Wide Web，简称 WWW）是一种信息查询系统，也是互联网上的一种服务。从某种程度来讲，由于它可以支持各类文本、图形、动画、声音，在互联网上越来越普及，几乎成为互联网的代称。

2. 超文本传输协议（HTTP）

超文本传输协议（Hyper Text Transfer Protocol，简称 HTTP）是客户端与服务器之间的通信协议，在 WWW 上使用。使用者通过这种协议向服务器索取资料，服务器也通过这种协议向客户端送回资料。

3. 统一资源定位符（URL）

统一资源定位符（Uniform Resource Locator，简称 URL）是每一信息资源在网上的唯一地址，由资源类型、存放资源的主机域名及资源文件名组成。例如：

`http: //www. interscience. wiley. com/index. htm/`

其中 `http` 表示资源类型，是超文本信息；`www. interscience. wiley. com` 表示主机域名；`index. htm` 表示资源文件名称。

URL 又称为信息资源的统一定位格式或统一资源定位器，实际上就是一个用以标识文档类型即其所在网络地址的字符串，它的用途是用统一的方式指明因特网上信息资源的位置。URL 包括三部分：所使用的传输协议、服务器名称、该服务器上定位文档的全路径名。URL 不仅用于 HTTP，还可用于 FTP、Gopher、Telnet 等协议。

只要用户正确地给出了一个文件的 URL 地址，WWW 服务器就能准确无误地将它找到并且传送到发出检索请求的 WWW 客户端上去。因此，URL 可以看成是一个文件在互联网上的标准通用地址。URL 的一般格式如下：

<通信协议> : // <主机> / <路径> / <文件名>

其中，<通信协议> 是指提供文件的服务器所使用的通信协议，如 WWW 的 HTTP 协议，Gopher 的 Gopher 协议，FTP 的 FTP 协议等；<主机> 是指上述服务器所在主机的 IP 地

址；<路径>是指该文件在上述主机的路径；<文件名>是指该文件的名称。例如，http://lib.nwpu.edu.cn/xxfw.htm/代表 WWW 上西北工业大学图书馆信息服务系统的 URL。

由此可见，通过 WWW 浏览器可提供多种互联网服务的访问。WWW 浏览器实际上为用户提供了一个统一的、一致性的交互接口，该接口利用 URL 中的 <通信协议> 信息来选择相应的客户程度，以访问相应的互联网上主机或服务器的信息。

4. 超文本文件 (HTML file)

HTML 是 Hype Text Market Language 的缩写，最大的特点是存在链接功能，可以从网页上的某一点直接指到另一个地方。

所谓超文本 (Hypertext)，就是非线性文本。一般人们阅读的文本 (text) 都是从上而下、从左到右排列，但在超文本中，内容是按超链接 (Hyperlink) 组织。用户单击文本中加以标注的一些特殊的关键词或图像，就能打开另一个文本。例如，用户在 WWW 上找到最新上映的新片介绍，点击感兴趣的链接，就可以连到有关的影评、演员剧照，甚至可以欣赏精彩片段；同时点击相关链接，还可以访问出售该片海报、唱片和其他宣传品的网上超市等。用户根本无须知道信息存储的具体位置，只需轻轻点击鼠标，按照事先安排好的链接，一层层地浏览、查询下去。而超媒体 (Hypermedia) 又进一步扩展了超文本所链接的信息类型，利用超链接将超文本和各种媒体信息连接在一起。用户不仅能从一个文本跳到另一个文本，而且可以激活一段声音，显示一个图形，或播放一段视频图像。在互联网中，每个 Web 服务器不仅提供其自身拥有的信息数据，还利用超链接指向其他的拥有相关信息的 Web 服务器，而这些服务器又指向更多服务器，通过这种内部的链接机制，使遍布全球的主机形成了一个相互联系、资源共享的有机整体。

5. 可扩展的标记语言文件 (XML file)

XML 是 Extensible Market Language 的缩写，是一种能够为各个特定领域创建标记语言的元标记语言 (meta - Language)，描述了文档内容的结构和语义，是建立在 HTML 的基础之上的。对用户来说，XML 文档表面上与 HTML 文档没有什么实质的差别，因此用户略知一二即可。

6. 主页 (home page)

主页是在与任何一个 WWW 服务器链接后所见到的第一页 (起始页)。它是该服务器入口处的 HTML 文件，是进入该服务器访问其所提供的各类信息资源的引导页。

所谓主页 (home page)，从表面上理解，就是某个单位、学校、企业，甚至政府、城市、国家在互联网上为自己建立起来的门面。人们从互联网访问这些地方的网站，首先都会接触到这个门面，并根据它的引导进一步查询该网站上的有关内容。用户首先在键盘上输入一个 IP 地址，接着系统响应其访问请求，并通过网络将对方的主页信息传递到用户的计算机上，这时屏幕上出现的通常是经过精心设计的图形界面，就是通常所说的主页。按照微软公司的比喻，如果把 WWW 当做是互联网上的大型图书馆，则每个站点就是一本，每个 Web 页面就是书的一页，主页则是书的封面和目录。用户可以从主页开始，通过 Web 链接访问各类信息资源，在 WWW 世界中漫游。

第二节 网络信息的产生与发展

网络是当今世界最大的信息资源宝库，人们通过网络可以了解到政治、经济、文化、教育、科技等诸多领域的最新信息。而网络上的信息又是浩瀚的，要想快速、准确地获取网络信息，就必须具备一定的信息素养，如信息检索能力、信息处理能力等。而要想得到这些能力，首先要对网络信息有一个清楚的认识。

一、网络信息的含义

信息广泛存在于自然界和人类社会之中。信息反映自然界和人类活动的状态、结果，任何事物皆有信息。然而，人们对信息的认知还要追溯到 20 世纪。1948 年，信息论的创始人申农（C. E. Shannon）在他发表的《通信中的数学原理》一文中认为，“信息是用来消除不确定的东西”。他的看法对后人影响重大。

网络信息的产生有两个重要条件，即互联网的地位提升和 Web 网页的作用日益重要。这两者作为网络信息的载体，对网络信息的产生和发展起着至关重要的作用。那么，什么是网络信息呢。朱静芳在《现代信息检索实用教程》一书中指出，“网络信息资源是指以电子数据的形式将文字、图像、声音、动画等多种形式的信息存放在光、磁等非印刷质的载体中，并通过网络通信、计算机或终端等方式再现出来的信息资源的总和”。可以看出，网络信息资源需满足两个条件：一是存储在光、磁等介质中，而不是存储在印本中；二是其检索和利用需要借助一定的工具，如通信工具、计算机及其他终端设备。

综上所述，网络信息就是以光、磁等形态存在，以互联网为介质传递，通过计算机等设备获取和利用的文字、图像、声音等信息的集合。

二、网络信息产生的背景

网络信息的产生是随着计算机的发明和互联网的出现而产生的，计算机为其提供了存储和利用的设备，而互联网则是网络传输的主要工具。

1. 计算机的产生与发展

电子计算机又称计算机或电脑，是一种利用电子学原理根据一系列指令来对数据进行处理机器。计算机的英文原词“computer”是指从事数据计算的人，而他们往往都需要借助某些机械计算设备或模拟计算机。这些早期计算设备的祖先包括算盘，以及可以追溯到公元前 87 年的被古希腊人用于计算行星移动的安提基特拉机器。随着中世纪末期欧洲数学与工程学的再次繁荣，1623 年德国博学家 Wilhelm Schickard 率先研制出了欧洲第一部计算设备，这是一个能进行 6 位以内数加减法，并能通过铃声输出答案的“计算钟”，使用齿轮来进行操作。

1941 年 5 月 12 日，德国工程师 Konrad Zuse 完成了他的图灵完全机电一体计算机“Z3”，这是第一部具有自动二进制数学计算特色以及可行的编程功能的计算机，但还不是“电子”计算机。

第二次世界大战中，为训练轰炸机飞行员，美国海军委托麻省理工学院实施旋风工程，制造出了世界上第一台能够实时处理资料的“旋风电子计算机”埃尼阿克，并发明

了磁芯存储器。这为个人电子计算机的发展做出了历史性的贡献。开发埃尼阿克的小组针对其缺陷又进一步完善了设计，并最终呈现出今天我们所熟知的冯·诺伊曼结构（程序存储体系结构）。这个体系是当今所有计算机的基础。

在整个 20 世纪 50 年代，真空管计算机居于统治地位。1958 年 9 月 12 日，在后来英特尔公司的创始人 Robert Noyce 的领导下，发明了集成电路，不久又推出了微处理器。1959 年到 1964 年间设计的计算机一般被称为第二代计算机。

到了 60 年代，晶体管计算机将其取而代之。晶体管体积更小，速度更快，价格更加低廉，性能更加可靠，这使得它们可以被商品化生产。1964 年到 1972 年间设计的计算机一般被称为第三代计算机。

到了 70 年代，集成电路技术的引入极大地降低了计算机生产成本，计算机也从此开始走向千家万户。1972 年以后的计算机习惯上被称为第四代计算机。1972 年 4 月 1 日，英特尔公司推出 8008 微处理器。1976 年，史蒂夫·乔布斯（Stephen Jobs）和斯蒂夫·沃兹尼亚克（Stephen Wozniak）创办苹果计算机公司，并推出其 Apple I 型计算机。1977 年 5 月，Apple II 型计算机发布。1979 年 6 月 1 日，英特尔发布了 8 位元的 8088 微处理器。

1982 年，微电子计算机开始普及，大量进入学校和家庭。1982 年 1 月 Commodore 64 计算机发布，价格 595 美元。1982 年 2 月 80286 发布，其时钟频率提高到 20 MHz，并增加了保护模式，可访问 16 M 内存，支持 1GB 以上的虚拟内存，每秒执行 270 万条指令，集成了 134000 个晶体管。

1990 年 11 月，微软公司发布第一代 MPC（Multimedia PC，多媒体个人电子计算机标准）：处理器至少为 80286/12 MHz（后来增加到 80386SX/16 MHz），有光驱，传输速率不少于 150 KB/s。1994 年 10 月 10 日英特尔公司发布 75 MHz Pentium 处理器。1995 年 11 月 1 日，Pentium Pro 发布，主频可达 200 MHz，每秒钟完成 4.4 亿条指令，集成了 550 万个晶体管。1997 年 1 月 8 日，英特尔公司发布 Pentium MMX，对游戏和多媒体功能进行了增强。

此后，计算机的变化日新月异，1965 年发表的摩尔定律不断被印证，预测在未来 10 ~ 15 年仍然适用。

2. 互联网的产生与发展

互联网（Internet）是指在 ARPA 网基础上发展起来的世界上最大的全球性互连网络。

20 世纪 50 年代，通信研究者认识到需要允许在不同计算机用户和通信网络之间进行常规的通信。这促进了分散网络、排队论和分组交换的研究。1960 年，美国国防部国防前沿研究项目署（ARPA）出于冷战考虑创建的 ARPA 网引发了技术进步并使其成为互联网发展的中心。1973 年 ARPA 网扩展成互联网，第一批接入的有英国和挪威计算机。

1974 年，ARPA 的鲍勃·凯恩和斯坦福的温登·泽夫提出 TCP/IP 协议，定义了在互联网网络之间传送报文的方法。1983 年 1 月 1 日，ARPA 网将其网络核心协议由 NCP 改为 TCP/IP 协议。

1986 年，美国国家科学基金会（National Science Foundation, NSF）创建了大学之间互连的骨干网络 NSFnet，这是互联网历史上重要的一步。1994 年，NSFnet 转为商业运营。1995 年，网络向商业开放，此时成功接入互联网的其他比较重要的网络还包括 Usenet、Bitnet 等多种商用网络。

1996年,“Internet”(互联网)一词被广泛流传。到此,以ARPA为基础的全球互联网络正式形成。

“Internet”又音译为“因特网”或“英特网”,在中国内地译作“互联网”,港澳地区也译作“互连网”。

第三节 网络信息的类型与特征

网络信息纷繁复杂、数量庞大,在方便人们学习、工作和生活的同时,也给用户带来了不便。这是由于网络信息类型多样、特征各异,如果不对它们进行分类整理,就不能很好地利用这些资源,甚至会造成资源越多越没有资源可用的恶性局面。本节将对网络信息的类型和特征加以阐述,以达到区分和鉴别的目的。

一、网络信息的类型

网络信息类型的划分有多个标准,可按照网络传输方式、信息整理的层次、信息的组织形式划分。

1. 按网络传输方式划分

1) WWW 信息

WWW信息即World Wide Web信息,它采用超文本传输协议(HTTP)进行传输,是目前互联网上使用最广泛的信息存储格式。WWW信息建立在超文本和多媒体基础之上,传输文本、声音、图像等,并以网页的形式反映出来。它是网络信息的主要部分,现在已成为最广泛使用的信息资源。

2) FTP 信息

FTP是互联网上常见的一种信息存取方式,它借助文件传输协议(File Transfer Protocol,FTP)在各计算机之间传输信息。用户可以通过它把自己的计算机与世界各地所有运行FTP的服务器相连,访问服务器上的大量程序和信息。FTP采用“客户端/服务器”方式,客户端要在自己的本地计算机上安装FTP客户程序才能使用它的资源。FTP信息目前仍是互联网上比较重要的信息。

3) Gopher 信息

Gopher(The Internet Gopher Protocol)信息查找系统,是互联网上一个非常有名的信息查找系统,它将互联网上的文件组织成某种索引,很方便地将用户从互联网的一处带到另一处,允许用户使用层叠结构的菜单与文件,以发现和检索信息。

Gopher是Web出现前互联网上最流行的分布式信息资源体系,互联网上曾有几千个Gopher服务器,它们遍布于大学、公司、研究机构或其他机构中。它具有简单、易用等特点,曾经非常受用户欢迎,但由于Web的出现,其地位一落千丈,现在所有的Gopher服务器都已被Web取代。

4) Telnet 信息

Telnet(Telecommunication Network Protocol)信息指在网络通信协议的支持下用户可以远程登录计算机,使自己的计算机在授权的情况下使用远程计算机的各种开放式的资源,如硬件资源和软件资源等。它是有效的实现远程连接计算机的方式,Diglog、OCLC、

DataStar 仍提供 Telnet 检索方式。Web 没有出现前, Telnet 是许多图书馆检索联机公共检索目录 (OPAC) 的主要方式。

2. 按信息的整理层次划分

1) 联机数据库 (database)

联机数据库是借助互联网传输, 并将其作为检索平台的一种存有信息资源的集中地。联机数据库中可存储文本资源、声像资源和其他多媒体资源等。联机数据库是目前最重要的网络信息之一, 其一般设有多种检索手段, 方便用户检索利用信息。数据库的主要类型有面向对象式数据库、数据仓库和分布式数据库。中文的数据库有万方数据库、维普数据库和 CNKI 数据库, 外文的数据库有 elsevier、springer、OCLC 及 BioMed 等。

2) 电子出版物

电子出版物是指以电子 (数字) 形式存在, 通过互联网传输, 包含有文字、图像、声音和视频等各种信息形式的出版物, 包括电子图书、电子连续出版物 (电子期刊、电子报纸)。计算机和互联网的出现给出版业带来了新的生机, 电子出版比传统出版节约资源, 是绿色的出版形式。电子出版物通过网络传输, 用户足不出户就可以享受到这些信息, 从而更受到用户的青睐。目前, 著名的报纸《华盛顿邮报》已经纯电子出版了, 其他报纸的印刷版数量也大大减少, 电子出版已是大势所趋。

3) BBS 论坛、博客、网络日志

BBS 论坛、博客、网络日志这三类都是以个人发布的形式出现, 其信息量大, 内容丰富, 信息时效性高, 但信息价值良莠不齐, 有时甚至有虚假信息, 需要加以选择甄别。

4) 其他信息

其他信息主要是个人和各类机构发布的信息, 这类信息量大而杂, 若想有效利用需进行分类整理。如新闻、各类机构的政策公告、市场交易、统计资料、广告、股票证券、评论等信息。

3. 按信息的组织形式划分

1) 文本信息

文本信息主要指静态的信息, 如 Web 网页信息、FTP 信息等, 可以有文本、图像、声音、视频等多种形式。文本信息包含的内容丰富、信息量大, 但其本身对于海量的信息有一定局限性, 不适合数量庞大的信息存储。随着用户信息素养的提高, 这种信息的形式越来越不适应用户需求, 只能作为一种辅助的信息组织方式出现。

2) 超媒体信息

超媒体是超文本与多媒体二者结合所产生的, 是二者在信息浏览环境下结合的跨媒体平台, 是 Web2.0 与全球化 3.0 即个人全球化、媒体化的有机聚合。如 E-mail、即时通信工具和博客, 都是超媒体的代表。这类信息形式发展迅速, 前景广阔, 是未来重要的信息资源类型。

3) 数据库信息

数据库技术是大量的数据和信息的规范管理和控制技术, 在这种技术背景下, 用户通过检索工具和检索语言, 可以轻松地查找到所需的信息。一般来说, 数据库存储的信息量都非常大, 且数据库中的信息规范化程度高、信息质量高, 特别是 Web 网页出现后, 数据库与其相结合形成网络数据库, 更加方便了用户利用, 也提高了数据库的价值。

二、网络信息的特征

前面已经说过，网络信息也是信息的一种，所以网络信息的特征也符合信息的特征，即网络信息具有普遍性、客观性、时效性、可共享性、可传输性、载体依附性等；同时，网络信息的类型有很多，而不同类型的信息其特征各异。总的来说，网络信息的特征主要体现在以下几个方面：

(1) 网络信息是普遍存在的，并且分布广泛，内容丰富。在互联网高速发展的今天，任何一个可以联网的通信设备或计算机都可以获得海量的、种类繁多的网络信息。用户可以通过这些信息了解全球各地实时发生的新闻，做到足不出户便能尽晓天下。

(2) 网络信息的类型多样。这里讲的类型多样，不但是指网络信息本身有多种类型，如文本形式、图像形式、声音形式等，还指传输网络信息的工具的类型多样，电子计算机、手机等都可以作为传输网络信息的工具。

(3) 网络信息分布广，发布的标准不统一，在小范围（如某个具体网站）有序，但在整个网络信息环境中处于无序状态。

(4) 获取网络信息的途径多，用户所花费的成本较低。除了数据库需要购买，数据库中的信息需要付费才能查看外，大部分网络信息是不收费或很少收费的。而且，很多网络信息的平台设有检索工具，用户可以很方便、快捷地检索并利用网络信息。

(5) 网络信息是开放式的，支持用户自由发表。博客、BBS、网络日志和 Wiki，都是用户自己上传的信息，这种网络信息是用户交流而产生的，更符合用户的需求，是发表自己的观点、见解的理想的园地。

(6) 网络信息的时效性很强，但很多信息不能长期保存。要查最新的新闻，网络无疑是个很好的选择，但网络信息在快速传递最新信息的同时，也会不自觉地剔除旧的信息，这就不利于回溯检索。这也是网络信息最大的一个缺陷。

总之，网络信息代表了新的信息类型，其发展潜力巨大，但缺点也比较明显，如果能解决不能长期保存的问题，那么网络信息将很可能发展迅速，以至于逐步取代传统的信息传播手段。

第四节 网络信息的分布

从总量来看，网络信息是数量巨大、内容十分丰富的，但如果按照具体的标准来细分，则网络信息的分布并不均匀，不但不同省份的网络信息质和量有差别，而且不同行业的网络信息也会有所不同，甚至同一省份同一行业不同时间的信息也有所差别。造成这种差别可能是地域发展的原因，可能是行业状况的原因，也可能是信息类型不同的原因。要彻底懂得网络信息的差异，就需要了解网络信息的分布情况。

一、网络信息的地域分布

网络信息的地域分布由于受到地区差异，如经济差异、文化差异和科技差异等的影响，分布很不均匀，具体来讲是东多西少，东部沿海省份比西部内陆省份的网站、网络数据库以及域名的数量要多。