

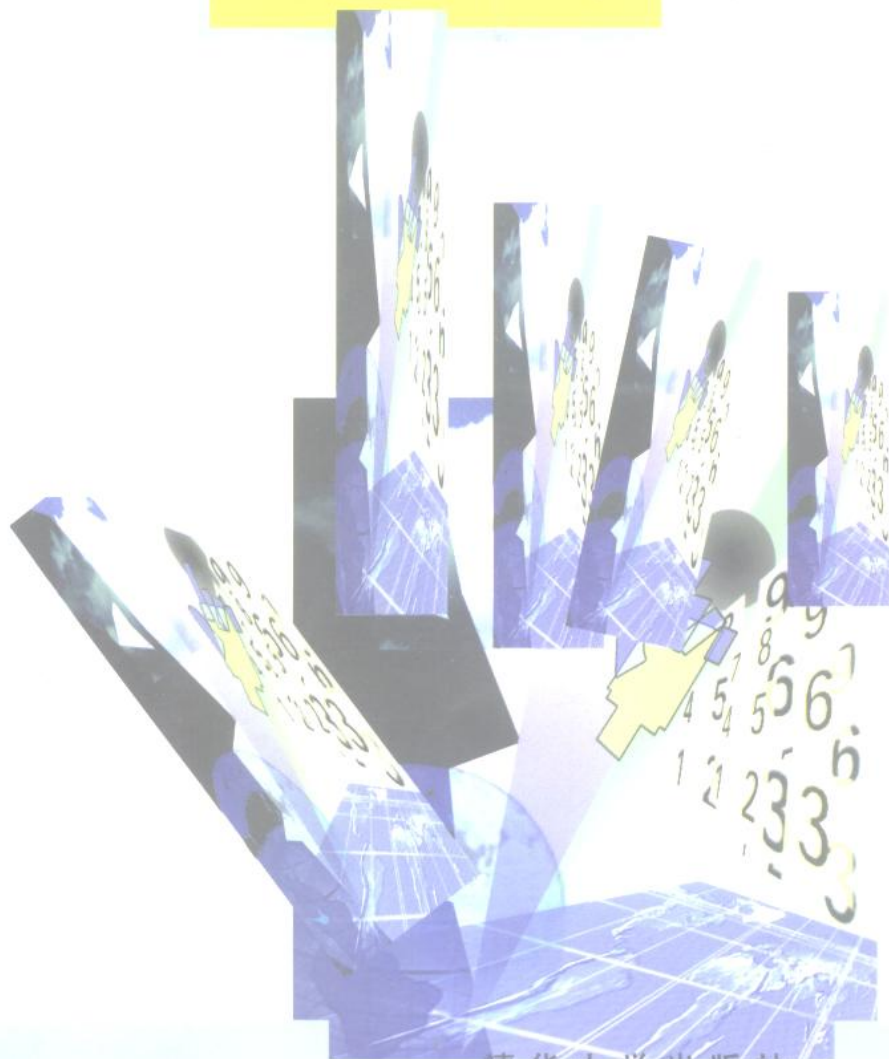
HTML

语言

Web

站点开发技术

马洪兵 张秋玲 著



清华大学出版社

<http://www.tup.tsinghua.edu.cn>



目 录

第一章 Internet 与 World Wide Web	1
1.1 Interent 的发展及主要服务	1
1.1.1 Internet 的起源与发展	1
1.1.2 TCP/IP 网络体系结构	3
1.1.3 IP 地址和域名系统	4
1.1.4 Internet 的应用服务	6
1.1.5 Intranet 简介	7
1.2 World Wide Web	8
1.2.1 World Wide Web 的产生与发展	8
1.2.2 World Wide Web 的特点	9
1.2.3 World Wide Web 的使用方式	10
1.2.4 在 Web 上发布信息的方法	11
1.3 HTML 语言	13
1.3.1 HTML 语言的演变	14
1.3.2 HTML 文档的编辑工具	16
第二章 HTML 语言的基本概念	19
2.1 HTML 语言与 SGML 语言	19
2.1.1 定义 HTML 语言的 SGML 文档	20
2.1.2 HTML 的词法	21
2.1.3 HTML 的注释	24
2.2 HTML 文档的特点与结构	25
2.3 HTML 3.2 标记一览	28
2.4 HTML 文档与 Internet 媒体类型	31
2.5 超链接	33
2.5.1 统一资源定位器 URL	33
2.5.2 超链接与锚点	34
第三章 文档单元与页首单元	36
3.1 HTML 标记的语法规则	36
3.2 文档单元	38
3.2.1 HTML	38

3.2.2	HEAD	39
3.2.3	BODY	39
3.3	页首节中的单元	41
3.3.1	TITLE	42
3.3.2	ISINDEX	43
3.3.3	BASE	44
3.3.4	STYLE	44
3.3.5	SCRIPT	46
3.3.6	LINK	47
3.3.7	META	48
第四章	字块级单元	52
4.1	标题	52
4.2	段落	54
4.3	预排格式文本	55
4.4	列表	57
4.4.1	无序列表	58
4.4.2	有序列表	59
4.4.3	定义列表	61
4.4.4	目录列表和菜单项列表	63
4.5	其他字块级单元	64
4.5.1	地址信息	64
4.5.2	水平标尺线	65
4.5.3	分区	67
4.5.4	居中对齐分区	67
4.5.5	块引用	68
第五章	文本级单元	70
5.1	字体风格单元	70
5.2	短语单元	73
5.3	锚点与超链接	78
5.3.1	锚点单元	78
5.3.2	链接到 HTML 文档	80
5.3.3	链接到外部媒体	82
5.3.4	链接到其他 Internet 服务	83
5.4	字体大小与颜色	85
5.5	内联图象	88
5.6	换行	92
5.7	客户端图象映射	94
5.8	Java 小程序	97
第六章	表格	101

6.1	表格的组成	102
6.2	描述表格的 HTML 标记	103
6.2.1	TABLE 单元	103
6.2.2	CAPTION 单元	104
6.2.3	TR 单元	104
6.2.4	TD 单元和 TH 单元	105
6.3	表格的控制	109
6.3.1	控制表格的表现方式	109
6.3.2	空表元	114
6.3.3	跨越多行多列的表元	116
6.3.4	在表格中嵌入图象或超链接	120
第七章	表单与 CGI 脚本	122
7.1	表单	122
7.2	表单内部的 HTML 单元	124
7.2.1	INPUT 单元	124
7.2.2	SELECT 单元和 OPTION 单元	132
7.2.3	TEXTAREA 单元	135
7.3	表单提交	137
7.4	CGI 的基本概念	139
7.4.1	CGI 脚本	140
7.4.2	CGI 脚本的输入与输出	140
7.4.3	CGI 脚本的编程	144
7.4.4	ISINDEX 标记	145
7.4.5	NPH 程序	148
7.5	表单内容的处理	149
第八章	Web 服务器	157
8.1	Web 服务器的工作原理	157
8.2	Web 服务器的安装与配置	160
8.3	服务器端包含	166
8.3.1	NCSA httpd 服务器的包含配置	167
8.3.2	Parsed-HTML 文档	168
8.3.3	服务器端包含的应用	170
8.4	日志文件	171
8.5	基于服务器的图象映射	172
8.6	服务器推进技术	179
8.7	基于 Windows 95/NT 系统的 Web 服务器	181
第九章	HTML 3.0 与非标准 HTML 扩展	187
9.1	HTML 3.0 草案	187
9.1.1	通用的属性	187

9.1.2 数学公式	189
9.1.3 插图	194
9.1.4 其他标记	195
9.2 Netscape HTML 扩展	197
9.2.1 嵌入对象	198
9.2.2 多帧窗口	198
9.2.3 其他 Netscape HTML 扩展	205
9.3 Microsoft HTML 扩展	208
9.3.1 对原有 HTML 标记的扩展	208
9.3.2 背景声音	213
9.3.3 MARQUEE	213
9.3.4 插入对象	215
第十章 Web 环境下的程序设计技术	219
10.1 Java 语言	219
10.1.1 Java 语言的特点	220
10.1.2 Java 语言的语法规则	224
10.1.3 Java 开发工具	236
10.1.4 Java 语言在 Web 中的应用	241
10.2 JavaScript	251
10.2.1 JavaScript 与 Java 的比较	252
10.2.2 JavaScript 的应用	253
10.2.3 JavaScript 的基本语法规则	254
10.2.4 JavaScript 中的对象	259
10.2.5 JavaScript 脚本的实例	265
10.3 ActiveX 和 VBScript	267
第十一章 Web 环境下的多媒体技术	272
11.1 Web 页面中的图象	272
11.1.1 图象格式	272
11.1.2 影响图象文件大小的因素	273
11.1.3 透明图象与交错图象	277
11.1.4 图象的获取与使用	278
11.2 Web 上的音频技术	280
11.2.1 声音信号的数字化	280
11.2.2 音频文件格式	281
11.2.3 在 HTML 文档中链接音频文件	282
11.2.4 音频流技术简介	282
11.3 Web 上的视频技术	284
11.4 Web 上的动画	288
11.4.1 GIF89a 动画技术	289

11.4.2 使用 Java applet 创建 Web 动画	291
第十二章 VRML 语言与 Web 上的三维世界	299
12.1 VRML 语言的起源与发展	299
12.2 VRML 的文档结构与坐标系	300
12.3 VRML 1.0 的节点类型	302
12.3.1 造型节点	302
12.3.2 点、线、面节点	303
12.3.3 文本节点	305
12.3.4 材料节点	306
12.3.5 纹理节点	310
12.3.6 变换节点	312
12.3.7 照相机节点	315
12.3.8 光线节点	316
12.3.9 组节点	317
12.3.10 锚点与内联节点	318
12.3.11 杂项	320
12.3.12 节点的定义与调用	320
12.4 VRML 场景的例子	321
12.5 访问 VRML 世界的方法	327
12.6 VRML 2.0 简介	328
第十三章 Web 信息服务系统的开发	330
13.1 软件工程的基本概念	330
13.1.1 软件生命周期模型	330
13.1.2 软件生命周期各阶段的主要任务	332
13.1.3 Web 信息服务系统的开发	333
13.2 系统分析	334
13.2.1 Web 信息服务系统的主题分析	334
13.2.2 信息内容分析	336
13.2.3 用户信息分析	337
13.2.4 软、硬件基础分析	337
13.2.5 开发成本分析	338
13.3 系统设计	340
13.3.1 设计方法	340
13.3.2 信息内容设计	341
13.3.3 链接设计与导航设计	345
13.3.4 Web 站点文件目录结构设计	350
13.3.5 页面布局与外观设计	351
13.3.6 主页设计	360
13.4 系统实现	363

13.4.1 系统实现小组的分工	364
13.4.2 Web 信息服务系统的测试	365
13.4.3 Web 信息服务系统的发布	367
13.5 Web 站点的运行与维护	369
附录 A HTML 文档创作工具	372
A.1 基于文本的 HTML 文档编辑器	372
A.2 所见即所得的 HTML 文档编辑器	379
附录 B HTML 3.2 文档类型定义	387
B.1 HTML 3.2 的 SGML 说明	387
B.2 HTML 3.2 文档类型定义	389
B.3 ISO Latin-1 字符实体	400
B.4 怎样阅读 HTML DTD	403
附录 C HTML 4.0 工作草案简介	407
参考文献	418

第一章

Internet 与 World Wide Web

Internet 是当今世界上最大的信息网络,它遍布全球,庞大无比,其规模和用户数量都是其他网络所无法比拟的。然而,只将 Internet 看成是一个计算机网络或者是互相连在一起的计算机网络的集合,是不能概括其含义的。Internet 不仅是一个网络,而且还是一个巨大的信息资源宝库,它所包含的信息从科研、教育到商业、艺术、娱乐,几乎无所不有。Internet 使人与人之间的信息交流更为方便及时,在这一意义上讲,Internet 已经把世界变得更小了。

然而,Internet 自诞生以来的大部分时间里却是默默无闻的,因为在 Internet 上进行信息交流与查询的传统方式,如 FTP、Telnet、E-mail 等,采用的都是命令方式,而且其命令十分繁琐难记。这样一来,Internet 在很长一段时间内只是教育科研界少数专业人员手中的工具,而几乎没有被广大普通用户所使用。

World Wide Web 的出现则改变了这一局面。World Wide Web 简称为 Web、WWW 或 W3,中文一般译为万维网、环球网等,本书中将 World Wide Web 简称为 Web。Web 为用户提供了美观友好的操作界面,并且提供了快速、方便、灵活的检索信息的方法。使用一种称为浏览器(browser)的软件,用户就可以访问过去要用 FTP、Telnet 等不同程序才能访问的信息资源。正是因为 Web 的简单易用,使它很快就风靡了整个世界,以至于在新一代的 Internet 用户看来,Internet 就等同于 Web。事实上,的确有很多人是先接触了 Web,才了解了 Internet,然后才开始使用 FTP、Telnet、E-mail 等 Internet 应用的。

本章作为全书的开篇,将简要介绍 Internet 的历史与现状,Web 的原理以及 HTML 语言的演变等基本内容,以便为后续各章提供必要的背景知识。

1.1 Internet 的发展及主要服务

Internet 是目前世界上最大的计算机互联网络,它从最初的设计开发发展到现在,已经成为了一个完整的、复杂的系统,其中涉及到许多技术,如体系结构、地址与域名管理、应用服务等,本节将简要介绍这些内容。在此之前,我们先来介绍 Internet 的发展历史。

1.1.1 Internet 的起源与发展

Internet 起源于 1969 年,当时美国国防部为了能在爆发核战争时保障军队内部的通信联络,由其下属的高级研究计划署(Advanced Research Projects Agency, ARPA)建立了一个由

4 台计算机互联而成的试验性的分组交换网络 ARPAnet。当时 ARPAnet 计划的目标是建立一个快速、方便的网络,使美国军方分布广泛、各自独立的计算机之间能够相互传输信息和数据,并且在诸如断电、线路中断、甚至遭受核打击等各种复杂的条件下,仍有能力自己处理和恢复故障,保证数据通信的稳定和可靠。

到了 1976 年,ARPAnet 上的节点计算机已发展到了 57 个,连接有各种不同类型的计算机 100 多台,联网用户达到 2 000 多人,其中包括了若干所大学中为美国军方工作的许多科研人员。在这种情况下,ARPA 开始把参与 ARPAnet 研究的科研人员召集到一起,开一些非正式的会议,共同探讨有关的技术问题。1979 年,ARPA 成立了一个非正式的委员会 ICCB (Internet Control and Configuration Board,网际控制与配置委员会),以协调和指导网际互连协议和体系结构的设计。新的网络协议定名为 TCP/IP,英文全称为 Transmission Control Protocol/Internet Protocol,即传输控制协议/网际协议。到了这时,Internet 一词才正式出现了。

1980 年,ARPA 开始把 ARPAnet 上运行的计算机转向新的 TCP/IP 协议。1982 年,美国国防部通过命令方式要求所有连入 ARPAnet 的网络必须采用 IP 协议(即 Internet 协议)互联。向 Internet 技术的转换于 1983 年初完成,这就是国际互联网为什么称为 Internet 的原因,而 ARPAnet 因此也成为了 Internet 的主干网。与此同时,美国国防部通信署(Defense Communication Agency,DCA)将 ARPAnet 分成了两个独立的网络,一个用于进一步的研究,仍然称为 ARPAnet,另一个则用于军事通信,称为 MILnet。

ARPAnet 网络和 TCP/IP 技术的成功,使美国国家科学基金会(National Science Foundation,NSF)认识到网络通信很快将成为科学研究的重要手段。为了使全国各科研教育机构的人员能够共享以前只供军方或为军方工作的少数科研人员使用的超级计算设施,NSF 于 1985 年出资在全美建立了五大超级计算中心,并且于 1986 年建立了一个新的称为 NSFnet 的高速信息网络。该网络互联了 NSF 的所有超级计算中心,并且全部连入了 ARPAnet,这样 NSFnet 就取代 ARPAnet 成为了 Internet 的主干网。NSFnet 采用的协议也是 TCP/IP。NSFnet 为 Internet 的推广做出了巨大的贡献,因为它面向全社会开放,从而就使 Internet 进入了以资源共享为中心的实用服务阶段。从此,Internet 得到了迅速发展,很快走向了整个世界。

Internet 在中国的发展也很迅速。1986 年,北京计算机应用技术研究所与德国 Karlsruhe 大学合作,启动了名为 CANET (Chinese Academic Network,中国学术网)的国际联网项目。1987 年 9 月在北京计算机应用技术研究所内正式建成了我国第一个 Internet 电子邮件节点,并于 1987 年 9 月 20 日 22 点 55 分,向世界发出了第一封电子邮件,标志着我国开始进入 Internet。1990 年,NCFC (National Computing and Networking Facility of China,中国国家计算机与网络设施)开始建设。NCFC 在国内也称为中关村网,它分为两层,低层为中国科学院院网(CASNET)、北京大学校园网(PUNET)和清华大学校园网(TUNET),高层为连接国内其他教育科研机构 and Internet 的 NCFC 主干网。1994 年 4 月,NCFC 正式开通了与 Internet 的专线连接,并于 1994 年 5 月 21 日完成了我国最高域名 CN 主服务器的设置,实现了真正的 TCP/IP 连接,可以向 NCFC 成员组织提供全部 Internet 服务。到了这时,我国才算正式进入了 Internet。

随后,我国又建成了 CERNET (China Education and Research Network,中国教育科研网),该网络的中心设在清华大学,目前已经连通了国内大多数高等院校和科研机构。1995 年 5

月,邮电部也开通了 CHINANET,这是一个面向全国的商业网,可以向全社会提供各种 Internet 服务。CHINANET 在国内与 CHINAPAC (中国公用分组交换网)、CHINADDN (中国公用数字数据网) 以及帧中继网和市话交换网等互联,从而向公众提供了多种连入 Internet 的方式。

今天,Internet 已经遍布世界各地,渗入到了社会生活的各个领域。目前到底有多少个网络以及有多少台计算机连入了 Internet,恐怕谁也说不清楚。可以说,Internet 在全世界的普及已经引发了一场划时代的变革,它促成了新的行业与服务的大量涌现,使人类社会迅速走向信息时代。

1.1.2 TCP/IP 网络体系结构

TCP/IP 是目前 Internet 网络体系结构和协议事实上的国际标准。与国际标准化组织 ISO 的 OSI (Open System Interconnection, 开放系统互连) 七层参考模型相比,TCP/IP 相对简单一些,共包含四个概念层:网络接口层 (Network Interface Layer)、网际层 (Internet Layer)、传输层 (Transport Layer) 和应用层 (Application Layer)。

网络接口层用于控制对本地局域网或广域网的访问,该层的协议有很多种,如以太网 (Ethernet)、令牌环 (Token Ring) 等。网际层负责解决计算机到计算机的通信问题,该层的协议主要为 IP 协议,IP 协议存在于主机和网关之中。传输层负责端到端的通信,即一个应用程序到另一个应用程序的通信。TCP 协议是该层的主要协议,它只存在于主机之中,用以提供可靠的数据传输。除了 TCP 协议之外,UDP 协议 (User Datagram Protocol, 用户数据报协议) 也是常用的传输层协议。因为更高层的应用层协议有可能利用 TCP 和 UDP,因此 Internet 上的主机绝大多数都支持 TCP 和 UDP,以扩大自己的应用范围。应用层包括若干网络应用程序,如电子邮件、查询服务等,该层的协议有很多,如 FTP 协议、SMTP 协议、HTTP 协议等。应用层的协议只在主机上实现。

TCP/IP 网络分层模型不十分符合 OSI 参考模型。大体上讲,网络接口层对应于 OSI 的物理层和数据链路层,网际层对应于 OSI 的网络层,传输层对应于 OSI 的传输层,应用层对应于 OSI 的会话层、表示层和应用层。TCP/IP 模型和 OSI 参考模型的对比如图 1.1 所示。

TCP/IP 模型	OSI 参考模型
应用层	应用层
	表示层
	会话层
传输层	传输层
网际层	网络层
网络接口层	数据链路层
	物理层

图 1.1 TCP/IP 模型和 OSI 参考模型的比较

TCP/IP 模型与 OSI 参考模型的一个重要区别是可靠性问题。OSI 模型在所有各层都进行差错检验和处理,而 TCP/IP 则将可靠性看作是端到端的问题,因此它只在传输层才解决差错检验和恢复问题,即允许单个链路丢失数据。

TCP/IP 的工作原理并不复杂。当文件在 Internet 上发送时, TCP 将文件分成小的 TCP 报文分组, 每个报文分组有一个控制头, 其中包括目的计算机地址、数据重组的信息和防止数据被破坏的信息。IP 则进一步把 TCP 报文分组分解为 IP 数据报, IP 数据报的控制头中有 IP 地址信息、装载的 TCP 信息和数据。

连续不断的 IP 数据报可能需要经过许多子网才能到达目的地。网络的交叉点设有路由器 (router), 它负责为 IP 数据报选择最佳传输路径, 以有效地分散 Internet 上的数据负荷, 避免在系统的某一部分造成拥塞。

一个文件的不同 IP 数据报可能要经过不同的路径到达同一地点。当所有 IP 数据报都到达目的地后, 首先去掉其中的 IP 地址信息, 然后利用 TCP 的相关信息检查数据在传输过程中是否受到了破坏。如果未发生破坏, 则由各数据报重新组成原文件; 如果发生了破坏, 则要求发送端重发被破坏的数据报。

这里要说明的是, TCP/IP 不是单个的协议, 而是一个很完整的协议簇。除了 TCP 和 IP 两个协议之外, TCP/IP 还包括其他协议, 其中有工具性协议、管理协议和应用协议等, 例如, HTTP、SMTP、UDP 等都是 TCP/IP 协议簇中的协议。所有 TCP/IP 协议簇中的协议都在 RFC (Request for Comments) 文件中予以规定和说明。

RFC 的含义是征求评论。RFC 文件是关于计算机通信领域的文档资料, 其内容从会议报告到协议说明书无所不包, 其长度从一二页到一本书厚薄不等。大多数 RFC 文件都是对网络协议与服务的描述和说明, 其中通常都给出了实现协议和使用服务的步骤细节。Internet 发展到今天的规模, RFC 的功绩是不可磨灭的。所有 RFC 文件的发表均有编号, 例如 RFC 1866 为 HTML 2.0 的规范, 而且一旦一个 RFC 文件以某个编号发表了, 则这个编号就不会再次使用了。Internet 工作委员会 (Internet Activities Board, IAB) 负责制定 TCP/IP 协议簇中的协议, 有关 TCP/IP 的 RFC 就是由 IAB 发表的。读者如果对 RFC 文件有兴趣, 可以访问 <http://ds.internic.net>; 此外, 从北京大学的匿名 FTP 服务器上也可以获取 RFC 文件, 其地址是 <ftp://ftp.pku.edu.cn/pub/rfc>。

TCP/IP 技术目前仍在不断地发展, 人们正在设计和开发下一代 TCP/IP 技术, 即 IP₆。目前 Internet 存在的主要问题是流量问题。随着入网成员的不断增加, 信息量急剧上升, 为此对高速通信技术提出了要求。近几年人们对 ATM (Asynchronous Transfer Mode, 异步传输模式) 技术的关注正说明了这一问题。

1.1.3 IP 地址和域名系统

Internet 上的每一台计算机都有一个唯一的地址, 称为 IP 地址。IP 地址共有 32 位, 可以写成 4 个用小数点分开的十进制数, 每个十进制数表示 IP 地址中的 8 个二进制位, 例如 IP 地址 1001 1000 0000 0010 0001 0110 0101 0001 可以写成 152. 2. 22. 81。

每个 IP 地址都由两部分组成, 分别为网络标识 (NetID) 和主机标识 (HostID)。网络标识代表一个网络, 而主机标识则代表该网络内的一台计算机。IP 地址分成 A、B、C 三类, 其格式如图 1.2 所示。A 类地址共有 $2^7 (= 128)$ 个, 均分配给大型网络使用, 每个网内的主机数可以达到 $2^{24} - 2$ (约为 1 678 万) 台; B 类地址共有 $2^{14} (= 16\,384)$ 个, 适用于中等规模的网络, 每个网内的主机数最多可以达到 $2^{16} - 2 (= 65\,534)$ 台; C 类地址最多, 有 2^{22} (约为 419 万) 个, 分配给小型网络, 每个网内的主机数最多为 $2^8 - 2 (= 254)$ 台。IP 地址可以描述主

Internet 上的各个网络依照 DNS 的命名规则对本网内的计算机命名。Internet 的最高层域名（顶级域名）由 Internet 协会（Internet Society）的授权机构负责管理。根据 Internet 国际特别委员会（IAHC）1997 年 2 月 4 日公布的关于通用顶级域名的报告，顶级域名分为三类：

Internet 上的各个网络依照 DNS 的命名规则对本网内的计算机命名。Internet 的最高层域名（顶级域名）由 Internet 协会（Internet Society）的授权机构负责管理。根据 Internet 国际特别委员会（IAHC）1997 年 2 月 4 日公布的关于通用顶级域名的报告，顶级域名分为三类：

- 由域名到 IP 地址的变换是由名字服务进行的。名字服务采用客户/服务器模式,名字服务器 (Name Server) 是一个运行在专门机器上的程序,提供从域名到 IP 地址的转换;名字服务系统的客户程序一般称为名字解析器 (Name Resolver)。名字服务提供的域名与 IP 地址的变换机制如图 1.3 所示。

由域名到 IP 地址的变换是由名字服务进行的。名字服务采用客户/服务器模式,名字服务器 (Name Server) 是一个运行在专门机器上的程序,提供从域名到 IP 地址的转换;名字服务系统的客户程序一般称为名字解析器 (Name Resolver)。名字服务提供的域名与 IP 地址的变换机制如图 1.3 所示。

例如,前面提到的域名 `ibm320h.phy.pku.edu.cn` 所对应的 IP 地址为 `162.105.160.189`。

1.1.4 Internet 的应用服务

由于目前 Internet 上应用服务的种类五花八门,而且不断有新的服务出现,因此这里只介绍几种最常用的服务。

1. 远程登录 (Telnet)

Telnet 是 TCP/IP 的一个应用层协议,即简单远程终端协议。它允许用户从本地计算机建立一条连接到某一远程服务器上的 TCP 连接,将本地计算机变为该远程机器的虚拟终端。

当用户利用 Telnet 登录到远程服务器上之后,本地计算机将完全成为该远程服务器的一个仿真终端。此时用户所能使用的功能和资源以及用户的工作方式将完全取决于该远程计算机系统。使用 Telnet 主要是为了共享某一远程计算机的软硬件资源和数据。

要在远程计算机上完成登录,则用户必须是该远程计算机系统的合法用户,即拥有该系统中相应的账户 (User ID) 和口令 (Password)。

2. 文件传输 (FTP)

文件传输服务是根据 FTP 协议命名的,FTP 是 TCP/IP 的一个应用层协议,称为文件传输协议 (File Transfer Protocol)。FTP 允许用户登录到远程服务器上,然后向远程机器传送文件或者从远程机器上获取文件,还可以远程执行少数简单的命令,如列出远程机器的文件目录等。

要想使用 FTP 存取远程服务器上的文件,则必须拥有该远程机器上的合法账户和口令,并拥有相应的访问权限。

匿名 FTP (anonymous FTP) 是一种特殊的 FTP 服务,它允许没有账户和口令的用户从远程计算机上获取文件。Internet 上有许多匿名 FTP 服务器,其上存放着大量的共享文件,包括软件、文档等。使用匿名 FTP 时,用户可以使用 `anonymous` 作为账户名,并使用自己的 E-mail 地址作为口令完成注册。

3. 电子邮件 (E-mail)

电子邮件 (Electronic Mail, 简称 E-mail) 就是通过计算机网络发送和接收邮件,它是一种快速、简便、高效、廉价的现代通信手段。很多人使用 Internet 主要就是为了收发电子邮件。据统计,Internet 上有一半以上的活动都与电子邮件有关。

电子邮件系统使用的是简单邮件传输协议 (Simple Mail Transfer Protocol, SMTP), 该协议主要负责解决如何通过一条链路把邮件从一台机器传送到另一台机器的问题。用户可以使用电子邮件的前提是在邮件服务器上拥有自己的邮箱。邮箱的地址称为 E-mail 地址,其形式为:用户名@邮件服务器域名,例如:

`mahb @ spark.phy.pku.edu.cn`

其中的符号 @ 表示的是英文单词 at (在,位于)。上述 E-mail 地址表示的是在域名为 `spark`。

phy.pku.edu.cn 的机器上的名为 mahb 的电子邮箱。

4. Gopher 服务

Gopher 是美国明尼苏达大学的研究人员创建的一种菜单界面的 Internet 信息检索服务。Gopher 本来是北美出产的一种地鼠的名字,以 Gopher 命名该服务,大概是意指该检索工具能像地鼠挖洞那样,一步步地达到目的地。另一种对于 Gopher 含义的解释是,Gopher 的英文发音与 go for (为你而去)的发音相同,意思是不论用户想查找什么,Gopher 都会为你而去。虽然这两种解释有所不同,但它们基本上都道出了 Gopher 服务的含义。

Gopher 可以将用户的请求自动转换成 FTP 或 Telnet 命令。它为用户提供了一级级的菜单,这些菜单动态地连接到 Internet 上的不同主机上,用户通过选取这些菜单就可以对 Internet 上的远程计算机系统进行访问。与 FTP 或 Telnet 等命令方式的信息服务相比,Gopher 的菜单方式对于用户是很方便的。Gopher 的缺陷在于它只能查询那些基于文本的信息资源。

据估计,目前在 Internet 上有数千个存储不同信息的 Gopher 服务器,其中绝大多数都是公开的,用户可以随意访问。

5. 网络新闻 (Network News)

网络新闻是建立在 Internet 平台之上的一种信息服务,也称为 Usenet,它表现为全世界范围内成千上万个网络新闻组 (newsgroup) 的集合。用户使用网络新闻服务可以与世界各地有共同爱好的人们就任何一个感兴趣的话题进行讨论,可以阅读网络新闻组中的信息或者在新闻组中发表自己的观点。用户加入或者退出一个新闻组都是完全自由的。

网络新闻系统目前存在 7 个大的专题组,它们是 comp (计算机)、news (新闻)、rec (娱乐)、sci (科技)、soc (社会)、talk (专题讨论) 和 misc (杂项)。

网络新闻服务采用的是网络新闻传输协议 (Network News Transfer Protocol, NNTP)。

6. World Wide Web

World Wide Web 是近几年最重要也是最热门的 Internet 应用,它采用的是超文本传输协议 (Hyper Text Transfer Protocol, HTTP)。由于我们很快就会专门介绍 World Wide Web,因此在这里就不多费笔墨了。

1.1.5 Intranet 简介

仅在三年以前,世界上还没有 Intranet 的概念,而现在,美国半数以上的大公司都已经建立了自己的 Intranet。在中国,也已有宝钢集团等大公司建立了企业内部的 Intranet。目前,Intranet 正在迅速增大,在 Netscape 公司的年营业收入中,已经有一半是来自于建立 Intranet 的客户了。

Intranet 是 Internal Internet 的缩写,一般译为内部网,它是利用 Internet 中的 TCP/IP 协议、Web 服务器、Web 浏览器、HTTP 协议等技术建立的供某一机构内部进行信息访问用的独立网络。Intranet 的含义分成两种:广义地讲,Intranet 是某一机构内部建立在 TCP/IP 协议之上的计算机网络;狭义地讲,Intranet 指的是机构内部使用的 World Wide Web。目前对于

Intranet 的理解基本上指的是后者。

Intranet 是机构内部的网络,为了保证其安全,一般采用防火墙 (firewall) 与 Internet 相隔离。所有由内部通向外部或者从外部通向内部的信息都必须经过防火墙,而防火墙则只允许那些得到了授权的信息从中通过,这样就保证了 Intranet 的安全性。

Intranet 只是对机构内部现有的网络设施做进一步的扩展,而不是抛弃原有的系统。机构中原有的内部网络只要在 TCP/IP 协议的基础上加装 Web 服务器、CGI、防火墙等,就可以建成 Intranet,并且与 Internet 相连。Intranet 可以充分利用原有的系统,因此建立方便并且花费低廉,这也正是 Intranet 得以迅速发展的一个原因。

1.2 World Wide Web

上一节中我们介绍了 Internet 上常用的几种服务,在这些服务中,除了 Gopher 和 World Wide Web 以外,其他的 service 都是基于命令行和字符方式的,即使是菜单界面的 Gopher 也只停留在字符方式的水平上。而 World Wide Web (以下简称为 Web) 的诞生,则使我们在 Internet 上获取信息的手段有了本质上的改善。有了 Web,即使一个不懂计算机网络的人也可以很快成为在 Internet 上查询信息的高手。可以说,Web 对于 Internet 的贡献正如 Windows 对于微型计算机的贡献一样重大,Web 的出现是 Internet 用户界面和使用方式的一次革命。

1.2.1 World Wide Web 的产生与发展

1990 年,World Wide Web 诞生于设在瑞士日内瓦的欧洲粒子物理研究中心 (Centre Europeen de Recherches Nucleaires, CERN),其主要研究人员是 CERN 的 Tim Berners-Lee。

关于核物理的研究分散在世界各地不同的国家进行,CERN 要通过 Internet 与世界各地的物理学家们进行学术交流。而当时 Internet 的使用方式是很不方便的,E-mail 只能发送文本信件,如果要发送其他类型的文件 (如图象),则要使用其他的 Internet 服务 (如 FTP)。为解决这一问题,Tim Berners-Lee 于 1989 年提出了一种新的构想,建议开发一种基于 Internet 的超文本 (hypertext) 系统,该系统应具有统一的用户界面,用户可以使用该系统查询各种类型的文件,如文本、图象等;此外,该系统还应具有跨平台的能力,即用户可以在 Internet 上使用各种不同类型的计算机进行信息查询。

基于 Tim Berners-Lee 的思想的 Web 服务器和 Web 浏览器于 1990 年在 CERN 开发成功,并开始在 CERN 内部广为应用。虽然当时的 Web 浏览器仍然是字符界面的,但是 Web 的思想无疑是先进的,因此 1991 年底 CERN 一经宣布 Web 服务,很快就引起了人们的重视。1992 年一年之间,全世界各地的 Web 服务器就增长到了 50 多台。

1993 年 2 月,位于伊利诺依大学分校的美国国家超级计算应用中心 (National Center for Supercomputing Applications, NCSA) 发布了一个 X-Window 系统下的图形用户界面的 Web 浏览器软件——Mosaic。Mosaic 一经推出即在全世界范围内引起了轰动,并使得 Web 迅速风靡了整个世界。因此,如果说 Web 的出现是 Internet 发展史上的一个里程碑,那么说 Mosaic 是 Web 发展中的一个里程碑是一点也不过分的。近两年在 Web 领域出尽风头的 Netscape 公司中,有很多开发人员当年都在 NCSA 开发过 Mosaic。

目前,Web 已经成为 Internet 上最为流行的信息查询手段。据估计,Web 上以字节计算的信息量目前正在以每年几万倍的高速度迅速增长。

1.2.2 World Wide Web 的特点

概括地讲,World Wide Web 是 Internet 上以超文本方式提供分布式、多平台超媒体信息服务的系统。

Web 是分布式的,因为 Web 上的信息分布在全球几百万台计算机上,只要是 Internet 连通的地方,就可以有 Web 信息。Web 用户可以通过 Web 浏览器访问世界各地计算机系统

中的信息。

当前主要的 Web 浏览器都支持各种操作平台,例如 Unix,Windows,Mac 等,因此 Web 是多平台的,用户可以在各种平台上访问 Web 信息资源。

超文本(hypertext)和超媒体(hypermedia)这两个词是美籍丹麦学者 T.Nelson 于 60 年代创造的。80 年代以后,随着个人计算机、计算机网络和多媒体技术的普及与发展,超媒体和超文本的概念与结构不断得到充实和完善,涌现出了一大批不同类型的超媒体、超文本应用系统,如 HyperCard、Toolbook 等等。

超文本是由节点(node)以及节点之间的链接(link)构成的语义网络,节点之间的链接也称为超链接(hyperlink)。超文本是非线性的网状结构,它提供了一种沿着链接访问数据的新方法,如图 1.4 所示。科学研究表明,人类的记忆是联想式的。构成一种网络互连结构;超文本使信息节点按照联想关系加以组织,其结构与人类的联想记忆结构类似。用户通过计算机阅读超文本时,可以看到一些显示方式与众不同的词语,这些词语一般带有下列线,并且采用一种更加醒目的颜色显示,这表明这些内容含有超链接。当用户对某个超链接感兴趣时,只需在该超链接上单击鼠标就可以跳到另一个超文本页面。在新显示出的超文本中,用户还可能发现新的感兴趣的超链接,于是可以继续沿着链接走下去。

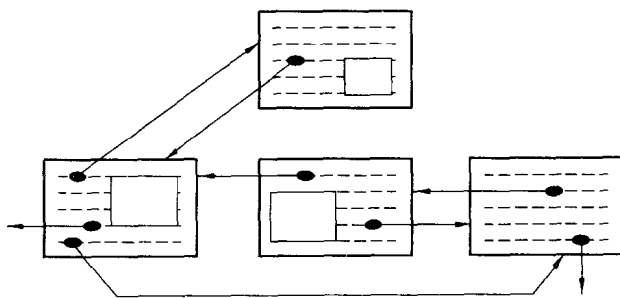


图 1.4 超文本

多媒体(multimedia)是 90 年代以来计算机领域中飞速发展的一项新技术,它使计算机表达与处理的信息从文字、数值扩展到图象、声音、视觉等多种形式。超媒体技术可以看成是超文本与多媒体技术的融合。在超媒体系统中,可以将多媒体信息,如声音、图象等结合到超文本中,而多媒体信息本身也可以像文字节点一样成为指向其他节点的超链接指针。因此,超媒体技术使超文本的内容更加丰富和生动,提高了表达思想的准确性和人机交互的友好性。

超文本和超媒体是密不可分的,而且实际上,在很多场合人们也已经不再区分超媒体和超文本了。超媒体的应用范围非常广泛,涉及文献管理、辅助教育、电子出版等多种领域。Windows 用户一定对联机帮助 (online help) 有较深的印象,而联机帮助其实就是超媒体与超文本技术的一种应用形式。

World Wide Web 也是一种超媒体系统,但与其他超媒体系统不同的是,Web 是基于 Internet 的,因此,Web 是一个跨越全球的超媒体网络系统,它把分散在全世界各地服务器上的文本、图象、声音和视频等信息资源有机地结合在一起,通过超链接在 Internet 上构成了一个巨大的逻辑网络,这也正是 World Wide Web (原意为遍布世界的蜘蛛网) 一词的含义和由来。

1.2.3 World Wide Web 的使用方式

World Wide Web 是典型的客户/服务器 (Client/Server) 体系结构。具体地讲,就是在存放信息的机器上运行一个称为 Web 服务器的软件,而在进行信息查询的用户机器上运行一个客户软件,即 Web 浏览器,如图 1.5 所示

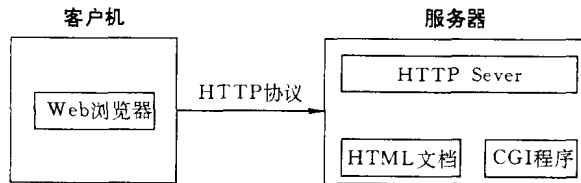


图 1.5 Web 的 Client/Server 结构模型

Web 服务器使用的主要协议是 HTTP 协议,HTTP 是为分布式超媒体信息系统设计的一个协议,最初于 1990 年提出,目前在 Web 上广泛采用的版本是 HTTP/1.0。HTTP 是一个无连接、无状态的协议,遵守客户/服务器体系结构。一个完整的 HTTP 事务由 4 步组成:

- 客户与服务器建立 TCP 连接
- 客户向服务器发送请求
- 如果请求被接受,服务器就向客户发送响应信息,其中包括状态码和所要的文件 (一般是 HTML 文档)
- 客户与服务器关闭连接

HTTP 协议支持的服务不仅限于 Web,它还可以支持其他服务,因此 HTTP 协议允许用户在统一的界面之下采用不同的协议访问不同的服务,如 FTP、Telnet 等。此外,HTTP 协议还可以用于名字服务。

在 Web 上,客户和服务器是一个相对的概念,只存在于一个特定的连接期间,在某个连接中的客户在另一个连接中也可能作为服务器。

目前可供人们选用的 Web 服务器 (有时也称为 HTTP 服务器) 软件有很多,比较著名的有 Unix 系统下的 NCSA httpd、CERN httpd 和 Windows 95/NT 系统下的 ZBServer、Website 等。

Web 浏览器的主要功能是解释并显示由 Web 服务器传送来的、由 HTML 语言写成的文档,包括嵌入在 HTML 文档中的 GIF 和 JPEG 格式的图象。此外,浏览器还可以根据用户的