

HTML 实用技术

裴有福 编著

万水Web技术精品丛书

2
/1



中国水利水电出版社
www.waterpub.com.cn

7p312
PYF/1

万水 Web 技术精品丛书

主编：裴有福

HTML 实用技术

裴有福 编著



中国水利水电出版社

www.waterpub.com.cn

050169

内 容 提 要

随着 Internet 的普及与发展,作为 Internet 重要服务的 Web 技术正在越来越深刻地影响人类社会的方方面面,超文本标记语言 HTML (HyperText Markup Language) 作为 Web 的核心,其地位已越来越重要,整个 Web 从一定意义上说就是构筑于 HTML 之上的,因此本书是每一个希望掌握最新计算机技术、特别是 Web 技术的人士的必读之作。

《HTML 实用技术》是 HTML 的参考工具书,全书共分十二章,系统地介绍了 HTML 的各个侧面,不仅覆盖 HTML 的基础、HTTP 协议、页面基础与进阶、页面多媒体、动态页面等 HTML 当前版本的全部内容,通过大量实例深入浅出地引导读者掌握 HTML,而且编入大量参考信息,同时还对 SSI 和最新的 HTML 4.0 进行了完整深入的介绍。本书适合 Web 技术人员使用,读者应对万维网 (WWW) 有一定的了解,具备基本的 Web 页面制作经验。

JS321/09

图书在版编目 (CIP) 数据

HTML 实用技术/裴有福编著. 北京: 中国水利水电出版社, 1998.6

(万水 Web 技术精品丛书)

ISBN 7-80124-739-6

I. H… II. 裴… III. 超文本标记语言, HTML, IV. TP312

中国版本图书馆 CIP 数据核字 (98) 第 16054 号

书 名	HTML 实用技术
作 者	裴有福 编著
出版、发行	中国水利水电出版社 (北京市三里河路 6 号 100044) 网址: www.waterpub.com.cn E-mail: sale@waterpub.com.cn 电话: (010)63202266(总机)、68331835(发行部)
发 行	新华书店北京发行所
经 售	全国各地新华书店
排 版	北京门槛创作室 WORD 照排部
印 刷	北京市天竺颖华印刷厂
规 格	787×1092 毫米 16 开本 17.5 印张 411 千字
版 次	1998 年 7 月第 1 版 1998 年 7 月北京第一次印刷
印 数	0001—5000 册
定 价	28.00 元

凡购买我社图书,如有缺页、倒页、脱页者,本社发行部负责调换

版权所有·翻版必究

前 言

不管您愿意不愿意、承认不承认, Internet 正在迅速改变着我们的生活习惯和生活质量, 并且必将越来越深刻地影响人类社会的方方面面。作为 Internet 的重要内容, World Wide Web (简称 Web, 缩写为 WWW, 中译为万维网) 已被公认为是 20 世纪的重大科技突破。它以一种简单方便、易于操作的方式链接着世界范围内各种类型的信息和各种层次的人们, 并将为人们构筑更加美好的生活和工作。在这种形势下, 学习和掌握 Web 技术, 就成为越来越多的人的迫切要求和强烈愿望。

《万水 Web 技术精品丛书》正是在这种形势下应运而生的。本套丛书的作者长期关注和研究 Web 技术, 目睹着它的迅猛发展, 也体会到它的意义重大。但同时也注意到许多 Web 使用者, 特别是一些初学者, 在广袤的 Web 空间里不知所措, 找不到重点, 分不清主次, 常常事倍功半。有很多读者, 没有条件或不能经常连入 Internet, 不能及时学习 Web 技术, 反被市面上的一些图书弄得懵懵懂懂。因此, 这套丛书力图系统全面地介绍 Web 技术的各个方面, 努力使读者能够在短期内迅速掌握 Web 技术。

超文本标记语言 HTML (HyperText Markup Language) 是 Web 的基础, 整个 Web 从一定意义上说就是构筑于 HTML 之上的, 它也正是 Web 获得成功的关键所在。虽然 HTML 并不是一种十分复杂的编程语言, 它的超文本思想简单地说就是纯文本加标记 (Tag), 但由于技术发展非常迅猛, HTML 处于不断发展和完善之中, 而且 HTML 也并没有简单到某些人所声称的那样, 不经学习即可制作 Web 页面的地步, 所以学习和掌握 HTML 仍然很有意义。

《HTML 实用技术》试图成为 HTML 的参考工具书, 它共分十二章, 全面系统地介绍了 HTML 的各个侧面, 不仅包括 HTML 当前版本的全部内容, 通过大量实例深入浅出地引导读者掌握 HTML, 而且编入大量参考信息和基础内容, 同时还对最新的 HTML 4.0 进行了完整深入的介绍。

衷心希望读者可以从本书中获得收益, 并在学习和掌握 Web 技术的过程中享受到无穷的乐趣! 本书作者努力作得更好, 但由于水平有限、时间紧迫, 不足与错误在所难免。希望读者和同行专家不吝指教, 以便在未来的版本中予以修改, 使本书真正成为 Web 技术的众望所归之作。



1998.5

目 录

前 言

第一章 HTML 基本概念	1
1.1 超文本及相关概念	1
1.1.1 超文本	1
1.1.2 超链接	2
1.1.3 页面与主页	2
1.2 浏览与搜索工具	3
1.2.1 Web 浏览器	3
1.2.2 Web 检索工具	5
1.3 URL	8
1.4 MIME	12
第二章 HTTP 协议	17
2.1 HTTP 概述	17
2.2 HTTP 的工作过程	19
2.3 HTTP 请求与响应	23
2.4 HTTP 头标	25
2.4.1 一般头标字段	26
2.4.2 请求头标字段	26
2.4.3 实体头标字段	27
2.4.4 响应头标字段	29
2.5 S-HTTP 与 SSL	29
第三章 HTML 概述	31
3.1 HTML 的基本语法	31
3.1.1 关于 SGML	31
3.1.2 HTML 字符实体	31
3.1.3 HTML 元素	32
3.1.4 HTML 元素属性	32
3.1.5 HTML 注释	33
3.2 HTML 重要约定和规定	34
3.2.1 HTML 文件的扩展名	34
3.2.2 HTML 文档的显示	34
3.2.3 HTML 文档字符集	35
3.2.4 HTML 数据类型	36
3.3 HTML 文档的总体结构	38
3.3.1 HTML 版本信息	38

3.3.2	HTML 元素	38
3.3.3	HEAD 元素	39
3.3.4	BODY 元素	41
第四章	HTML 开发工具	43
4.1	普通文本编辑器	43
4.2	专用 HTML 编辑软件	44
4.2.1	Netscape Composer	44
4.2.2	Microsoft FrontPage	45
4.2.3	HotDog Pro	47
4.2.4	NetObjects Fusion	49
4.3	HTML 文档转换工具	50
4.3.1	Microsoft Word 97	50
4.3.2	其他转换工具	51
4.4	HTML 辅助工具	51
4.4.1	Paint Shop Pro	51
4.4.2	PhotoImpact GIF Animator	52
4.4.3	GoldWave	53
第五章	页面基本技术	55
5.1	导引示例页面	55
5.2	页面总体结构	55
5.2.1	页面结构	55
5.2.2	示例页面的首次实现	56
5.3	文档字体设置	57
5.3.1	标题元素	58
5.3.2	FONT 元素	58
5.3.3	字符修饰元素	59
5.4	页面版式设置	61
5.4.1	分行控制	61
5.4.2	文本对齐方式	63
5.4.3	列表元素	65
5.4.4	预格式化文本	68
5.4.5	文档块处理	70
5.5	页面修饰	72
5.5.1	水平分隔线	72
5.5.2	页面分栏	73
5.5.3	闪烁文本	74
5.6	超链接	74
5.6.1	HREF 属性	74
5.6.2	NAME 属性	76

5.6.3	TITLE 属性	76
5.6.4	TARGET 属性	76
5.6.5	其他属性	77
5.7	文档信息增强元素	77
第六章	页面高级技术	79
6.1	表格 (Table)	79
6.1.1	表格的基本语法	79
6.1.2	跨多行、多列的表元	80
6.1.3	表格尺寸设置	80
6.1.4	表格对齐与布局	82
6.1.5	表格的标题	84
6.1.6	表格的颜色	85
6.1.7	结构化表格	86
6.1.8	表格线的显示	88
6.1.9	表格综合示例	89
6.2	表单 (Form)	94
6.2.1	表单基本用法	95
6.2.2	文字和密码输入	96
6.2.3	复选框和单选框	96
6.2.4	图像输入	97
6.2.5	隐藏表单组件	98
6.2.6	列表框	98
6.2.7	文本框	99
6.2.8	HTTP 文件上传	100
6.3	窗框 (Frame)	101
6.3.1	窗框的基本用法	101
6.3.2	窗框的排列与尺寸设置	102
6.3.3	窗框的互操作	103
6.3.4	窗框的外观	104
6.3.5	窗框设计示例	106
第七章	页面多媒体技术	108
7.1	图像 (Image)	108
7.1.1	基本用法	108
7.1.2	图像质量控制	109
7.1.3	图像对齐与布局	109
7.1.4	客户端图像映射	110
7.2	音频和视频	113
7.2.1	文档声音	113
7.2.2	多媒体文本	113

7.2.3	内联视频	113
7.2.4	VRML	114
7.3	滚动文本 (Marquee)	115
第八章	页面动态技术	118
8.1	页面动态技术概述	118
8.2	推送与拖曳技术	119
8.2.1	推送与拖曳的比较	119
8.2.2	服务器推送	120
8.2.3	客户端拖曳	122
8.3	嵌入脚本程序	123
8.3.1	JavaScript	123
8.3.2	VBScript	124
8.4	嵌入 Java 小应用	125
8.5	嵌入对象	127
8.5.1	EMBED 元素	127
8.5.2	OBJECT 元素	129
第九章	SSI	131
9.1	SSI 概述	131
9.2	在文档中插入信息	132
9.2.1	echo 标记	132
9.2.2	include 标记	133
9.2.3	fsize 标记	134
9.2.4	flastmod 标记	134
9.3	在文档中执行操作	134
9.3.1	exec 标记	134
9.3.2	config 标记	135
9.3.3	odbc 标记	136
9.3.4	email 标记	138
9.4	在文档中进行逻辑比较	139
9.4.1	if 标记	139
9.4.2	goto 标记	140
9.4.3	label 标记	141
9.4.4	break 标记	141
第十章	HTML 4.0 文档结构	142
10.1	HTML 4.0 概述	142
10.1.1	HTML 4.0 设计原则	142
10.1.2	HTML 4.0 文档设计原则	144
10.1.3	HTML 4.0 的几个基本概念	144
10.2	文档总体结构	145

10.2.1	HTML 版本信息	145
10.2.2	HTML 元素	145
10.2.3	HEAD 元素	145
10.2.4	BODY 元素	148
10.3	语言信息和文本方向	152
10.3.1	lang 属性: 指定内容的语言	153
10.3.2	dir 属性: 指定文本的方向	154
10.4	文本: 段落、行与字词	156
10.4.1	白空格 (white space)	156
10.4.2	结构化文本	157
10.4.3	分行和段落	159
10.4.4	标记文档的变化: INS 元素和 DEL 元素	161
10.5	列表: 无序列表、有序列表和定义列表	162
10.5.1	无序列表和有序列表: UL、OL 和 LI 元素	162
10.5.2	定义列表: DL、DT 和 DD 元素	163
10.5.3	DIR 元素和 MENU 元素	163
10.6	表格	164
10.6.1	表格结构	164
10.6.2	表格格式化	170
10.6.3	表格示例	172
10.7	链接: 超链接和媒体无关的链接	174
10.7.1	路径信息: BASE 元素	175
10.7.2	链接和锚的定义	175
10.7.3	A 元素	176
10.7.4	LINK 元素	178
10.8	包容: 对象、图像和小应用	180
10.8.1	对象包容: OBJECT 元素	180
10.8.2	图像包容: IMG 元素	185
10.8.3	小应用包容: APPLET 元素	186
10.8.4	文档包容	187
10.8.5	图像映射包容	188
10.8.6	图像、对象和小应用的可视化表述	190
10.8.7	替换文本的指定	190
第十一章	HTML 4.0 文档表述	192
11.1	样式表: 控制文档的表述	192
11.1.1	给 HTML 添加样式	193
11.1.2	继承与层叠	198
11.1.3	隐藏样式元素的内容	198
11.1.4	通过 HTTP 头标指定样式	199

11.2 对齐方式、字体样式和水平分隔线	199
11.2.1 格式化	199
11.2.2 字体	202
11.2.3 水平分隔线: HR 元素	203
11.3 窗框: 文档的多视图表述	203
11.3.1 窗框的布局	203
11.3.2 指定目标窗框信息	206
11.3.3 替换内容	207
11.3.4 内联窗框: IFRAME 元素	208
第十二章 HTML 4.0 交互文档	210
12.1 表单 (Form)	210
12.1.1 FORM 元素	210
12.1.2 控件	211
12.1.3 元素焦点的设置	220
12.1.4 失效元素和只读元素	222
12.1.5 表单提交	223
12.2 脚本 (Script)	223
12.2.1 脚本文档的设计	224
12.2.2 不支持脚本时的文档设计	229
附录 A HTML 元素和属性索引	231
附录 B 浏览器支持 HTML 的比较	237
附录 C HTML 字符集	244
C.1 ISO Latin-1 字符集	244
C.2 基本英文字符集	245
附录 D Netscape 定义的 HTML 颜色	253
附录 E HTML 4.0 对 HTML 3.2 的改进	255
E.1 对元素的改进	255
E.2 对表格的改进	255
E.3 对表单的改进	255
附录 F HTML 4.0 元素索引	257
附录 G HTML 4.0 属性索引	260

第一章 HTML 基本概念

World Wide Web 是非常博大精深的。它是从理论到实践的典范，其成功有着深厚的理论基础。超文本标记语言 HTML (HyperText Markup Language) 对 Web 的成功起到了举足轻重的作用。

本章将介绍与 HTML 有关的一些基本概念，主要为以下三个方面：与 HTML 页面有关的重要概念，统一资源定位符 URL 的介绍以及 MIME 的基础内容。

1.1 超文本及相关概念

Web 是信息发展与超文本技术的结合，它的信息访问是双向的：一方面是 Web 浏览，用户可通过浏览器在 Internet 上寻找所需的信息；另一方面是 Web 展示，用户可在自己的 Web 服务器上建立和存放按一定结构组织的超文本文件，供网络其他客户浏览，传播知识、交流信息和进行广告宣传。

Web 技术的基础是超文本技术，本节介绍超文本及有关的重要概念。

1.1.1 超文本

超文本 (HyperText) 是把一些信息根据需要链接起来的信息管理技术。人们可以通过超文本中的链指针打开另外一个相关的文本，用鼠标点击文本中的高亮度或带下划线的文字，即可打开与之相链接的文本，获得所需的信息。目前已有许多软件提供的联机帮助信息或联机手册都是基于超文本的。

超文本由若干互连的文本块组成，这些信息块可以是若干屏、窗口、文件或更小的块信息，这样的信息单元称为节点 (node)，如图 1-1 中的 A、B、C、D。不管节点有多大，每个节点都有若干指针指向其他节点或从其他节点指向该节点的指针，这些指针称为链接 (link)。超文本链接亦称为锚点 (anchor)，或称锚标。图 1-1 所示为一个文本、四个节点、五条链接组成的超文本结构。

构成超文本的信息网络的每个节点都包含一个特定主题的信息，节点的大小视主题而定。按照节点的功能主要可分为表现型节点、组织型节点和推理型节点三类。

表现型节点是一般性超文本节点，主要有：由文本段组成的文本节点；图形图像节点；由声音或合成语音构成的声音节点；用于执行一个过程的按钮节点。

组织型节点是组织节点的节点，主要有：目录节点，它以条目形式给出信息的索引指针，指向索引节点；索引节点，由索引项组成，索引项用指针指向目的节点或相关索引项。

推理型节点只有智能型的超文本中才有，它主要包括对象节点和规则节点。

超文本的“超”体现在它不仅能包含文本，而且可以包含图像、音频、视频等多媒体信息，即将文本的概念扩充为超文本。因此，超文本也称超媒体 (Hypermedia)。

超文本的书写格式有专门的规定，这就是超文本标记语言 HTML (HyperText Markup Language)。HTML 的版本变化非常快，已由 1.0 版发展到当前通用的 3.2 版，最近还发布了最新的 4.0 版，因而尚未成为国际标准。

在 Internet 上传输超文本文件必须遵循超文本传输协议 HTTP (HyperText Transfer Protocol)。超文本传输协议是 Web 的基本协议, 本书将在下章较为详尽地予以介绍。

文本 1

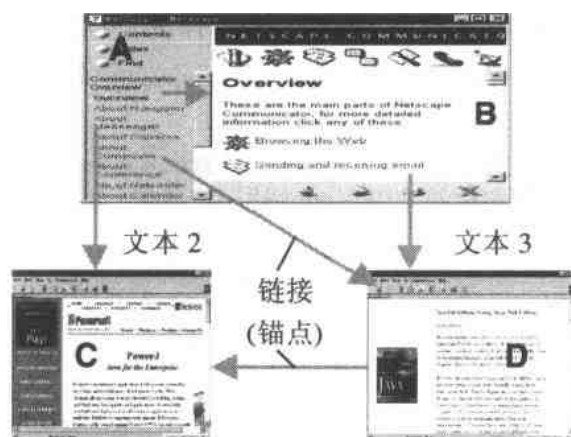


图 1-1 超文本及相关概念

1.1.2 超链接

链接 (link) 是超文本的灵魂, 它定义了超文本的结构。从图 1-1 中可以看到链接的示意。超文本中的链接主要包括以下几类:

- 基本链接: 用于建立节点之间的基本顺序, 类似文章的章、节、小节这样的层次结构。
- 索引链接: 可以从一个索引点引到该节点的索引入口。
- 执行链接: 将一种执行活动与按钮节点相连, 使超文本成为高层程序的界面。

在浏览器显示的超文本中, 链接表现为带下划线的彩色字符串或图案。如果链接被访问过, 显示颜色将改变, 以便提醒用户。链接的颜色还可通过 HTML 参数来改变。

锚点 (anchor) 是链接的指针, 它指向同一文档的其他部分或其他不同的文档。在 HTML 语言中有两种锚点:

- NAME 锚点: 它属于标记锚点, 用来将本文档中的某个或某些段落标记为节点, 供用户通过链接查找信息。
- HREF 锚点: 它属于执行锚点, 用来指向其他超文本节点, 或引用本文档中用 NAME 锚点标记了的段落, 或执行某个 URL 命令 (如发送电子邮件、执行 FTP 或 HTTP 命令)。因此, HREF 锚点又可分为 URL 命令锚点和内部链接锚点两种。

锚点元素语句在浏览器中显示时就是链接, 所以有一些 HTML 文献直接称呼锚点为链接, 即认为锚点就是链接。

通过 HREF 锚点执行的链接称为超链接 (Hyperlink)。不过, 在大多数情况下并没有严格区分链接与超链接, 例如 Web 页面中的链接也可以叫做超链接。

1.1.3 页面与主页

从用户角度看, Web 是由巨大的、遍及全球的电子文档组成的, 这些文档通常简称为

页面 (page)。每个页面都可以由一个或多个超文本文档组成, 其中包含指向其他页面的超链接。

一个 Web 站点通常包含大量页面, 这些页面的首页称作主页 (Home Page), 它是用户用浏览器进行浏览时默认看到的第一个页面。用 Web 页面展示信息时, 首先要设计主页, 它应该是画面精美的简要目录。

主机的文件名应该与 Web 服务器系统配置文件中指定的 Web 默认页的文件说明一致 (通常为 index.html 或 default.html 等), 以确保 Web 访问者连接到 Web 主机时可以直接看到主页。

Web 表述 (Web presentation) 是包含许多连接在一起的、用于表达某一目的的页面的总称。它是 Web 页面的体系结构。在设计 Web 页面时, 首先应该有一个通盘考虑, 将 Web 这种体系结构规划得更合理, 使其结构化, 便于将来进一步设计和扩展。



提示与技巧

页面或主页文件名的扩展名一般应该采用 .html。对于 UNIX 操作系统, 如果使用 .htm, 那么当网上的 UNIX 用户浏览时只能看到源文件, 而 Windows 用户浏览 .htm 和 .html 时都能看到正确的显示结果。因此 UNIX 下页面文件的扩展名一定要用 .html。对于 DOS 和 Windows 3.x 这类操作系统, 由于文件名受到“8.3”规则的限制, 其扩展名只能取作 .htm。Windows 95 和 Windows NT 则两者都可以。

1.2 浏览与搜索工具

Web 浏览器是 Internet 资源浏览软件, 被称作超媒体工具。它不仅可以显示多媒体的 Web 页面, 还可以通过 URL 命令连接到不同的 Internet 服务器上获取信息。

Web 导航工具则是在 Internet 上查找和获取信息的高效工具。

1.2.1 Web 浏览器

Web 浏览器是一种访问 Web 服务器的客户端工具软件, 使用它可以在非常友好的界面下方便地进入 Internet, 获取信息。

Web 浏览器的最基本目的在于让用户在自己的计算机上检索、查询、采掘、获取 Web 上的各种资源。由于 Internet 正处在日新月异的飞速发展阶段, Web 每天都被使用它的人们赋予新的涵义, 使得浏览器的功能也在不断的扩充和更新。目前市面上已有几十种浏览器, 功能有强有弱, 归纳起来, 浏览器应具备以下几种基本功能:

- 检索查询功能: 浏览器读入 HTML 文档, 解释 HTML 所描述的图表、声音、动画、表格, 以及进一步的链接信息, 利用 HTTP 协议, 可在任意 Web 服务器上畅游。

- 文件服务功能: 能在下载文档时实时查阅该文档, 并可利用 HTTP 去跟踪感兴趣的链接。当感到不需要正在下载的文档时, 可以随时中止下载过程。可对正在查阅的文档随时进行保存、打印、前后浏览等操作。

- 热表管理: 浏览器应能够自动记住用户刚刚访问过的 Web 地址, 称为“热表”。当用户想要回到刚才曾访问过的某一 Web 中, 用户可以在热表中快速地切换。

- 主页制作: 一些浏览器可以浏览 Web 页面, 而且也带有 HTML 编辑工具, 使用户

能够制作自己的 Web 页面。

- 其他 Internet 服务: 浏览器正在努力提供越来越多的 Internet 服务, 如 FTP、Gopher、WAIS、Telnet、Usenet 及 E-mail 等。

浏览器的功能可以通过配置得以扩充。对于浏览器不能处理的超媒体信息, 它检查配置文件, 查看如何处理接收到的信息。通常, 在配置文件中会给定外部查看器 (external viewer) 或帮助程序 (helper application) 的程序名, 浏览器自动调用这些程序处理自己不能处理的信息。如果某类文件没有指定外部查看器或帮助程序, 浏览器就询问用户对此文件的处理方法。通过这种方法还可以让浏览器调用其他 Internet 服务程序 (例如 TELNET), 从而扩展浏览器的功能。

目前市面上流传着数量非常多的浏览器, 它们大多为免费或共享软件, 可以从 Internet 上方便地获取 (如果是商业软件, 则通常可以下载其测试版或试用版)。最著名的浏览器包括:

1. NCSA Mosaic

早期的浏览器只能浏览 ASCII 文本文档。第一个基于图形且得到广泛使用的浏览器是 Mosaic, 该浏览器由美国国家超级计算应用中心 NCSA (National Center for Supercomputing Applications) 的软件开发部设计编写。Mosaic 是一种多用途的 Web 工具, 能够支持信息的定位、合并与恢复。

Mosaic 能够支持许多操作系统平台, 如 X-Window、MacOS 和 Windows 等。它也能够支持多种协议, 并允许用户用不同的方式显示文档。

Mosaic 不仅直接导致了后来极为著名的 Netscape 的产生, 而且也对其他浏览器产品产生了深远的影响。



背景与历史

在 Web 基本协议提出不久的 1993 年, Web 的发展引起了设在美国伊利诺伊州立大学的 NCSA 的 Marc Andreessen 和他的同事们的兴趣。他们迅速研制出一种软件, 能够解释 Web 中的 HTML 文档, 并将 HTML 文档包容的信息以统一的方式显示出来。这就是 Mosaic。这是第一个真正的 Web 浏览器, 最早运行在 Sun 工作站 X-Window 图形环境上。后来, NCSA 又推出了 Mosaic 的 Macintosh 及 Windows 版本。

2. Netscape Navigator

Netscape Navigator (常简称为 Netscape) 几乎已经成为浏览器的同义词。Netscape 包括了 Internet 应用的所有工具, 除浏览和制作 Web 外, 还可用于电子邮件、网络新闻、网络视频会议等。它支持各种多媒体对象, 包括多窗页面 (frame)、插件 (plug-in)、JavaScript 和 Java 小应用程序。它还能管理嵌入了带有音频、视频信息的三维对象的 Web 文档。

在 Netscape 的新版本 (4.0 版以后) 中, 还包括支持最新的推送 (Push) 技术的 Netcaster 组件, 以及信息与日程管理工具等。Netscape 正在向操作系统的方向进军。



背景与历史

1994 年 4 月, Mosaic 的主要设计者 Marc Andreessen 和 SGI 公司的主要创始人之一 James H. Clark 共同创办了 Netscape 通信公司, Marc Andreessen 任技术副总裁。在他的率领下, Netscape 通信公司在 Mosaic 的基础上进行了大量改进, 开发出了一种新型的浏览器——

Netscape Navigator。Navigator 提供了一个与 Mosaic 相似但比它更实用的图形界面。它的各方面性能均有了很大的增强，是第一个优先快速显示文本和图形的浏览器，也是第一个在收到整页文档前就允许查看首页的浏览器。

3. Microsoft Internet Explorer

Microsoft 曾经对 Internet 不够重视。不过，它起步虽晚，但有雄厚资金作后盾，很快就追了上来。Internet Explorer 是 Microsoft Internet 战略的重要产品。

目前，Microsoft Internet Explorer 已经成为功能强大的主要浏览器之一。它支持 Java、HTML 及其扩展、VRML、JavaScript 和 ActiveX 等。

4. Sun HotJava

Sun Microsystem 公司为了展示他们发明的 Java 语言，专门用 Java 开发了一种浏览器，称为 HotJava。HotJava 不仅继承了在 Mosaic、Netscape 等浏览器中实现的 Web 浏览技术，而且扩充了把静态文档变成动态可执行代码的动态特性。HotJava 显示了 Java 语言环境的强大功能，同时还为在 Internet 上发布 Java 程序提供了一个理想的平台。

5. WorldView

目前 Web 站点中显示的大部分页面是基于文本和二维图像的，但用虚拟现实造型语言 VRML (Virtual Reality Modeling Language) 定义三维对象的页面也越来越多。

以前许多浏览器都不支持 VRML，要浏览 VRML 对象必须借助专门的浏览器。WorldView 就是一个较为著名的 VRML 浏览器。

现在只要适当配置 Netscape 和 Microsoft Internet Explorer 等浏览器，就可以直接浏览 VRML 对象。



网址与站点

访问以上浏览器的站点，可以获得关于浏览器的最近信息和详细内容，还可以直接下载浏览器软件。它们的 URL 分别是：

Mosaic: <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic>

Netscape: <http://www.netscape.com>

Microsoft Internet Explorer: <http://www.microsoft.com>

HotJava: <http://java.sun.com>

WorldView: <http://www.intervista.com/products/worldview/index.html>

其他的浏览器也很多，其中较为知名的包括：Spry 公司开发的 AIR Mosaic、Spyglass 公司的 Enhanced NCSA Mosaic、NaviSoft 公司的 GNN Works、BookLink Technologies 公司的 Internet Works、Quarterdeck 公司的 Quarterdeck Mosaic、IBM 内置于 OS/2 Warp 内的 Web Explorer 以及 Macintosh 上使用的 MacWeb 等。

1.2.2 Web 检索工具

Web 所具有的超文本和超媒体结构，带来了信息出版和传播的一场革命。但是，由于 Internet 是连接大量网络和计算机的巨型国际网络，其上的信息资源极为丰富，加上网络发展不均、网上信息良莠不齐及其他技术和非技术因素，人们在浏览 Web 的时候，遇到的最大困难就是如何迅速、准确地从浩如烟海的信息资源中找到自己最需要的信息，这已经成

为困扰全球网络用户的最主要问题之一。

面对此种情况, 计算机专家和信息管理专家正在积极探索和研究, 采取各种方式方法解决这一困难。从目前技术发展现状看, Web 用户可以通过三种方式检索信息:

1. 使用浏览器直接浏览

这是最普通、最传统的信息获得方式。用户根据文件中的链接追踪信息的发展脉络, 从一个页面跳到另一个页面。这是一种很有趣的网中冲浪, 但比较费时, 检索费用大, 而且容易迷失方向。

2. 按照主题指南进行查询

在 Web 中, 人们已经编制了各种主题指南, 提供可查询和检索的层次式主题目录, 以超链接方式将不同学科、专业、行业和区域的信息按照分类或主题目录的方式组织起来。用户通过主题目录的指引, 逐渐找到与自己需求有关的信息。



网址与站点

目前 Web 上最著名的主题指南包括:

- 英国图书馆布告板协议主题目录 (BUBL): <http://www.bubl.bath.ac.uk/BUBL/Tree.html>。它以国际十进制分类法和英文字母顺序组织学术机构和学术出版物, 对从事学术研究的人员有很高的参考价值。

- WWW 协会编辑的虚拟图书馆 (The Virtual Library): <http://www.w3.org/hypertext/DataSource/bySubject/Overview.html>。它集中了大量学科领域的页面, 还参照美国国会图书馆主题词表对页面加以组织。

- 美国密西根大学开发的主题资源指南信息中心 (The Clearinghouse for Subject-Oriented Resource Guides): <http://www.lib.umich.edu/chhome.html>。它适合于查询商业和学术信息。

3. 利用搜索引擎 (Search Engine) 进行关键词、主题词或自然语言查询

搜索引擎是目前检索 WWW、Gopher 和 FTP 文件最主要的工具软件。将搜索引擎用于网络资源的管理和检索是 Web 上检索信息最主要的发展趋势。除英文搜索引擎外, 多语种和其他语种的检索工具也不断出现。目前全球已开发成功的检索工具软件不下 200 种, 对推动网络信息的使用和传输作出了重要贡献。

搜索引擎由三个部分组成:

(1) 巡视软件 (英文称作 robot、knowbot、knowledge robot、wanderer、crawler、worm 或 spider 等)。巡视软件穿行于 Web 空间, 在页面间自动追寻超链接, 从服务器上采集新出现的信息, 并对信息进行标引, 确定页面之间的链接是否有效。巡视软件一般先访问知名度较高的服务器, 并鼓励 Web 用户递交自己的地址, 以便发现和收录新的信息。

(2) 数据库。巡视软件将采集和标引的信息记录到数据库中。不同软件以不同方式进行记录, 有的记录页面全文, 有的仅收录地址、篇名、题名、特定的段落或重要的词。不同搜索引擎的数据库规模大小各异, 更新方式也不相同。

(3) 检索代理 (Agent)。当用户提出查询要求时, 由查询软件代替用户在数据库中进行检索, 并将结果返还给用户。在这一过程中, 该软件还会利用特殊算法对文件和用户的需求进行计算和评估, 以便更好地满足用户的需求。

不同搜索引擎在访问 Web 服务器时，可能只访问软件开发者指定的站点（选择性引擎），也可以访问任何一个服务器（非选择性引擎）；在标引信息时，有的按广度优先原则进行（只追踪服务器顶端的文件目录），有的按深度优先原则进行（深入到服务器文件内部，并跟踪其超链接）。

目前已投入使用的搜索引擎可分为独立式引擎和集成式引擎两大类。独立式引擎使用单独的检索工具，其数据库可以是全文数据库或非全文数据库。集成式引擎汇集若干引擎的使用界面，用户可将查询交付不同引擎进行检索。

表 1-1 列举了目前较为知名的搜索引擎。



网址与站点

表 1-1 中各搜索引擎的 URL 分别为：

Alta Vista: <http://altavisa.digital.com>

Excite: <http://www.excite.com>

OpenText: <http://www.opentext.com>

Lycos: <http://www.lycos.com>

WebCrawler: <http://webcrawler.com>

AliWeb: <http://www.traveller.com/aliweb>

WWWorm: <http://www.cs.colorado.edu/home/mcbryan/WWW.html>

Infoseek: <http://www.infoseek.com>

Yahoo: <http://www.yahoo.com>

CUI W3: <http://cuiwww.unig.ch/W3catalog>

表 1-1 著名 Web 搜索引擎的简单情况

类型	搜索引擎	引擎开发者	数据库范围
全文数据库类	Alta Vista	DEC 公司研究实验室	URL、HTTP、FTP 及新闻组。1996 年底数据库中已包括 3000 万个 URL
	Excite	Architext 公司	5000 万个页面、3 万个 Web 站点，包括评论和分类广告
	OpenText	Waterloo、Ontario 与 UUNET	HTTP、FTP 和新闻组。已包括 2000 万页面以上
非全文数据库	Lycos	卡内基·梅隆大学机器翻译中心	HTTP、FTP、Gopher 及图像、音频、视频文件。提供评论等增值服务
	WebCrawler	华盛顿大学的 Brian Pinkerton	HTTP、Gopher 和 FTP。22 万已标引的页面和 3600 万未经标引的页面
	AliWeb	英国 Xerox 公司	服务器管理者参与数据库建立。规模较小，但质量较高