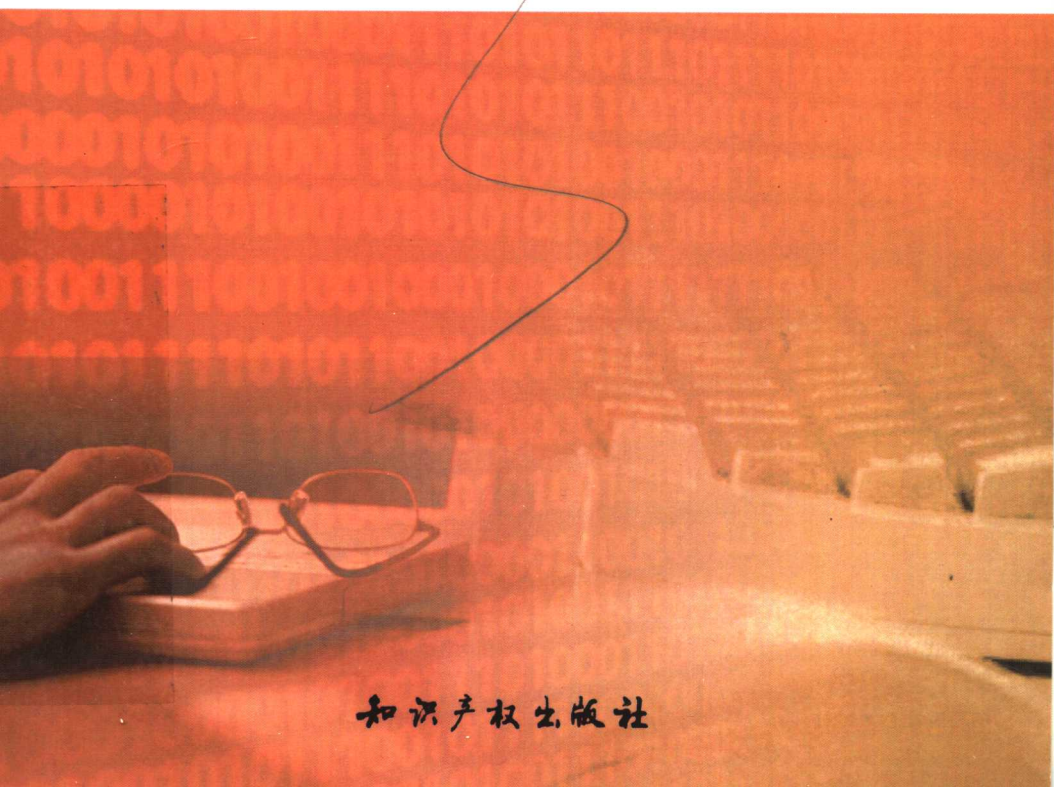


Web XINXI JISHU JIAOCHENG

潘有能 叶 鹰 / 编著

Web 信息技术教程



知识产权出版社

Web 2.0 时代网络应用案例教程

第 2 版 杨 勇 主编

Web 信息技术教程



清华大学出版社

TP393.4/200

2007

Web 信息技术教程

潘有能 叶 鹰 编著

知识产权出版社

内容提要

本书汇集了制作和发布 Web 信息的主要技术与应用,包括 HTML 知识、XML 知识、网页制作工具软件、搜索引擎技术、ASP 交互技术、服务器技术等。全书理论与实践相结合,融入了作者的经验和见解,可作为 Web 制作与发布的技术教程或技术指南,适合用作 Web 网页和网站制作类课程或培训班教材,也可供有一定经验的 Web 网络管理员和有关技术人员参考。

责任编辑:蔡虹

图书在版编目(CIP)数据

Web 信息技术教程/潘有能,叶鹰编著. —北京:知识产权出版社,2007.6
ISBN 978-7-80198-621-4

I. W… II. ①潘…②叶… III. 计算机网络—应用程序—教材
IV. TP393.09

中国版本图书馆 CIP 数据核字(2007)第 094443 号

Web 信息技术教程

潘有能 叶鹰 编著

出版发行:知识产权出版社

社址:北京市海淀区马甸南村 1 号

网 址: <http://www.ipph.cn>

发行电话:010-82000893 82000860 转 8101

责编电话:010-82000860 转 8324

印 刷:知识产权出版社电子制印中心

开 本:880mm × 1230mm 1/32

版 次:2007 年 6 月第 1 版

字 数:300 千字

ISBN 978-7-80198-621-4/T · 249

邮 编:100088

邮 箱:bjb@cnipr.com

传 真:010-82000893

责编邮箱:caihong@cnipr.com

经 销:新华书店及相关销售网点

印 张:10.625

印 次:2007 年 6 月第 1 次印刷

定 价:30.00 元

版权所有 侵权必究

如有印装质量问题,本社负责调换。

前 言

Web 全称为 World Wide Web，也简称为 WWW，翻译为万维网，是建立在 Internet 上的一种应用。1991 年，Web 由在欧洲粒子物理研究所工作的 Tim Berners - Lee 发明（Tim Berners - Lee 因此当之无愧地获得 2004 年由芬兰技术奖励基金会授予的 100 万欧元“千年技术奖”）。Web 与统一资源定位器（URL）、浏览器（Browser）曾一起被合称为 Internet 上的三大发明。今天，Web 已成为 Internet 的最大功能群，揽括了 80% 以上的 Internet 信息流量，与 E - mail、FTP 并称为 Internet 上的三大应用。

在 Web 技术飞速发展的同时，Web 信息的制作与发布正在成为一种新兴职业——Web 网络管理员（Webmaster）。目前国内外已经出版了不少关于 Web 制作方面的书，但大多偏重某一方面，如 HTML、FrontPage、Java 等，涉及 Web 信息创作与发布全面技术的不多。本书全面汇粹 Web 信息创作与发布的应用技术，可望为有志于成为 Web 网络管理员的读者提供一部体系完整的参考书。

显然，Web 正在形成继印刷书刊和广播电视后的第三次媒体和传播革命，Web 技术必将深刻影响各行各业的运营模式，甚至社会生活的各个方面，因而了解和掌握 Web 技术意义重大。

本书由潘有能博士和叶鹰博士合作编写，潘有能起草第 2、3、5 章和 6.1 节，叶鹰起草第 1、4 章和 6.2 节，书中实例多是作者自作或改编的，具有简单实用的特点，同时也具有较好的源码完整性。

目 录

1 Web 信息组织: HTML	(1)
1.1 HTML 的基本结构	(1)
1.2 HTML 的基本技术	(4)
1.3 HTML 的高级技术	(34)
2 Web 信息集成: XML	(82)
2.1 XML 概述	(82)
2.2 XML 文档的结构与语法规则	(89)
2.3 XML 的显示	(102)
2.4 XML 链接	(115)
2.5 应用程序接口	(124)
3 Web 制作工具: Macromedia Studio MX	(134)
3.1 Dreamweaver	(134)
3.2 Fireworks	(182)
3.3 Flash	(200)
4 Web 信息检索: 搜索引擎	(216)
4.1 搜索引擎技术原理	(217)
4.2 搜索引擎的检索方法	(221)
4.3 万维网搜索引擎 (Web Search Engines)	(224)
4.4 元搜索引擎 (Meta - Search Engines)	(240)
4.5 专用搜索引擎	(246)
4.6 搜索引擎技术的未来趋势	(254)
5 Web 交互技术: ASP	(257)
5.1 ASP 简介	(257)

Web 信息技术教程

5.2	ASP 脚本语言: VBScript	(262)
5.3	ASP 内置对象	(275)
5.4	ASP 中数据库的操作	(290)
6	Web 服务器技术	(308)
6.1	Windows 2000/XP Web 服务器: IIS	(308)
6.2	UNIX 服务器	(322)
	参考文献	(333)

1 Web 信息组织: HTML

HTML 是 HyperText Markup Language (超文本置标语言) 的缩写, 它起源于 SGML (Standard General Markup Language, 标准通用置标语言), 是 SGML 的一个子集。HTML 是目前组织 Web 页面信息的主要技术。有人以为现在已有不少主页制作工具便可以不必学习 HTML, 这是错误的, 工具软件主要是对业余爱好者有利, 对于专业 Web 网络管理员或深入研习者来说还是需要了解 HTML 的, 这样才能更有效地制作和组织网页。

HTML 是一种类似于排版用途的标记性语言, 用 HTML 语言编写出的 Web 网页 (以 .htm 或 .html 为后缀), 其实质是文本文件, 用户可以使用任何文本编辑器来进行编制。

HTML 的标准联机资源可以参考 <http://www.w3.org/Markup/>。

其他相关联机资源可以参考 HTML Help (<http://www.htmlhelp.com>)、HTML Writers Guide (<http://www.hwg.org>) 等。

1.1 HTML 的基本结构

HTML 文档中包含有大量的标记 (Tag), 如 < HTML >、< HEAD >、< BODY > 等。这些标记只是描述网页内容在浏览器中显示的方式, 它是由尖括号括起来的, 绝大多数是成对出现的, 前面的称为头标, 如 < BODY >, 表示在浏览器显示时对后面的内容开始用该标记方式; 后面的称为尾标, 如 < /BODY >, 是在头标记前加 “/” 并放在 < > 中, 表示该标记的作用结束。也有小部分 HTML 标

记是单个出现的，做一些特殊的操作，如 <HR> 标记表示在浏览器上画一条水平线，
 标记表示回车换行。标记本身在浏览器中并不显示出来。

标记 (Tag) 是组织 HTML 文件的基本符号。大多数标记都由起始标记 (Starting - tag) 和结束标记 (Ending - tag) 构成，其一般形式为：

<TAGNAME> ... </TAGNAME>

或 <TAGNAME ATTRIBUTENAME = value> ... </TAGNAME>

或 <TAGNAME>

标记名 TAGNAME 是标记的标识。任何一个标记都只有一个名字，如 HEAD、BODY 等。

标记有自己的属性，它是对标记更详细的补充，属性有相应的属性值。例如， 标记有 SRC 属性， 则表明所插入的图片的文件名及路径。

有的标记会有几个可用的属性。如果同时使用两个属性时，则中间可用空格分开。

HTML 文档应该置于 <HTML> 和 </HTML> 标记之间，换言之，文档的开头应该是 <HTML> 命令，最末尾应该是 </HTML> 结束命令，使浏览器和服务器能将文档标志成一个 HTML 文档。

紧接着 <HTML> 命令便是文档头的开始，文档头以 <HEAD> 命令开始并以 </HEAD> 命令结束，在文档头命令之间可以加入 <TITLE> 命令和 </TITLE> 终结命令，置于 <TITLE> 和 </TITLE> 命令之间的一切将作为该页面的标题，显示在 Web 浏览器的标题栏上，当访问者要将该页面做成书签 (Book Mark) 时，标题即为书签的名字。

HTML 文档的主要部分是文档体，它以 <BODY> 命令开始，以 </BODY> 终结，将显示在浏览器的正文区域中。

任何一个 HTML 文档至少由两部分组成：页面的头和页面的主体，页面头包含页面的版本说明、标题等，页面主体包含了所需要编

辑的页面的内容。因此, HTML 文档也就定义了相应的头标记 `< HEAD > ... </HEAD >` 和体标记 `< BODY > ... </BODY >`。下面就是一个 HTML 文档的基本结构。

<code>< HTML ></code>	(表示 HTML 文档定义开始)
<code>< HEAD ></code>	(表示头定义开始)
<code>< TITLE ></code>	(表示标题开始)
WEB	(标题内容)
<code></TITLE ></code>	(表示标题结束)
<code></HEAD ></code>	(表示头定义结束)
<code>< BODY ></code>	(表示页面主体开始)
MAIN DOCUMENT INFORMATION	(主体内容)
<code></BODY ></code>	(表示页面主体结束)
<code></HTML ></code>	(表示 HTML 文档结束)

通常, 编写 HTML 时常写成压缩形式如下。

```

< HTML >
< HEAD > < TITLE > WEB </TITLE > </HEAD >
< BODY >
    MAIN DOCUMENT INFORMATION
</BODY >
</HTML >

```

这两种形式除了在 Internet 上传输的字节数多少不同外并没有什么区别。

利用 HTML 语言开发 Web 页面时应注意:

1) HTML 语言不区分 `< >` 中的大小写。

2) HTML 语言忽略空白字符, 如空格、回车、换行、制表符等(在 `< PRE > ... </PRE >` 中使用的除外), 即几个空格的作用与一个空格的作用是一样的。回车换行用 `< BR >` 标记代替, 注意 `< BR >` 命

令一般用在行尾，它没有形如 `</BR>` 的命令结束标记。分段用 `<P>` 代替，建议在段尾使用 `</P>`，当然也可以不使用。

3) 对于显示控制码的字符，如 `<`、`>`、`&` 等，需要其他方式来显示，如 `<` 代表 `<`，`>` 代表 `>`，`&` 代表 `&`，`"` 代表双引号等。

4) 关于 HTML 文档中的注释，同其他语言一样，HTML 语言也允许用户在 Web 页面中加入注释。通常，注释内容加入方法是：

`< !... 注释内容 ... >`

一种特殊的注释是 `< DOCTYPE >` 标记，它指明使用哪一个 HTML 版本和遵守特定的 HTML 版本。例如，`< ! DOCTYPE " HTML 3.2//EN" >`，表明遵循 HTML 3.2 版。

1.2 HTML 的基本技术

除上述 `< HTML >`、`< HEAD >`、`< BODY >` 等结构性标记外，HTML 的基本技术是通过块标记（Block Tags）和文本层标记（Text - level Tags）来实现的，前者用于安排段落和规范页面布局，后者用于定义字体风格和超链接元素等。

1.2.1 块标记（Tags）

块标记包括以下几项。

- 1) 标题：从 `< H1 > ... </H1>` 到 `< H6 > ... </H6>`。
- 2) 预定格式显示：`< PRE > ... </PRE>`。
- 3) 分段：`< P > ... </P>`。
- 4) 回车：`< BR >`。
- 5) 水平分隔：`< HR >`。
- 6) 文本块：`< DIV > ... </DIV>`、`< CENTER > ... </CENTER>` 等。
- 7) 引用：`< BLOCKQUOTE > ... </BLOCKQUOTE>`。

8) 地址: <ADDRESS>... </ADDRESS>。

9) 列表: ... 、... 、<DL>... </DL>。

10) 表格、表单、框架等高级元素。

实际上,要在浏览器中显示一段信息,最简单的处理方法就是在<BODY>... </BODY>内用<PRE>... </PRE>标记将这段信息括起来并以文本格式存盘,这特别适用于端点信息的处理。使用该标记使框中标记中的内容在浏览器上显示时保持原来的格式显示,保留原文的空白符,如回车、换行等。

在<PRE>中也可以定义信息显示的宽度,这一功能通过<PRE>的一个属性来定义,如:<PRE WIDTH="80">会使浏览器每行显示80个字符。WIDTH的默认值为80,一般建议使用40、80、132这几个宽度。

在<PRE>... </PRE>中,还可以加入字体、字号等标记,也可以使用行和段落标记<P>、
等。

1.2.1.1 标题

在文章中为了表达清楚同时又突出重点,常常会采用章、节、段的层次结构,即以不同大小的标题来划分文章的结构。在HTML语言中用于定义标题大小的标记为<Hn>,其中n为标题的等级。HTML文档中的标题等级分为6级,具体规定见表1.1。最大的字号为<H1>,最小的字号为<H6>。

和命令用来使文本看上去是更粗或更黑的显示,<I>和</I>则用于使文本看上去是更细或更淡的斜体显示。因此,可以在想以黑体显示以示强调的文本开始处加上,并以结束。也可以将和命令嵌套于其他一切HTML命令中(例如,在<H1>和</H1>命令中间)。

表 1.1 标题的格式

标题的标记符	意义
<code><H1>... </H1></code>	产生一个黑体、居中的标题，使用非常大的字体，自动紧跟一个空行或两个空行，总是打印在新一页的顶部，常作为文件的开头
<code><H2>... </H2></code>	产生一个黑体、左对齐的标题，使用大号字体，前后各空一或两行
<code><H3>... </H3></code>	产生一个斜体、左空白边缘稍缩进的标题，使用大号字体，前后各空一或两行
<code><H4>... </H4></code>	产生一个黑体、比 H3 左空白边缘缩进更多的标题，使用正常字体，前后各空一行
<code><H5>... </H5></code>	产生一个斜体、比 H4 左空白边缘缩进同样大小的标题，使用正常字体，前面空一行
<code><H6>... </H6></code>	产生一个黑体、左空白边缘缩进的标题，使用正常字体，前面空一行

此外，需强调的文字还可用 `<EM>`、`</EM>` 标记（斜体），更加强强调则用 `<STRONG>`、`</STRONG>` 标记（黑体）。

“字体大小”的标记和“指定字体”的标记可以组合使用，例如：

```
<I> <FONT SIZE=5>
    <B> 今天 </B> 天气 <FONT SIZE=3> 真好! </
FONT>
</FONT> </I>
```

浏览器显示为 6 个斜体字，其中强调“今天”为黑体，5 号字；“天气”为通常的 5 号字；“真好!”为通常的 3 号字，显示如下：

今天 天气 真好!

需要闪烁、引用和地址时, 处理方法如下:

闪烁用 `< BLINK > ... </BLINK >` 标记, `< BLINK > ... </BLINK >` 内的信息在浏览器中显示时会自动闪烁。

引用采用 `< BLOCKQUOTE >` 标记, `< BLOCKQUOTE >` 会自动缩进相对于其他主体文本的页边空白, `< BLOCKQUOTE >` 标记能使你在安排引用文字时使用。

地址采用 `< ADDRESS >` 标记, `< ADDRESS >` 标记可插入一个斜体地址。

1.2.1.2 段落编排

`< P >` 和 `< BR >` 标记被用来标志段落和行控制, 这些标记只是在文本里插入回车; 注意 `< BR >` 没有相应的终结标记 `</BR >`。`< BR >` 标记仅仅将文本移到下一行的开头, `< P >` 标记作相同的工作, 同时加入一回车。它们在段首一般不产生一个左缩进, 而是顶头显示。一般地, 用 `< P >` 标记分开的同一段落的文字根据页面宽度自动换行, 但也有些浏览器不能自动换行。由 `< P >` 产生的段与段之间的空白比 `< BR >` 换行产生的要大, 可看作插入一空行。举例如下。

THIS IS `< BR >` AN EXAMPLE `< P >` OF THE `< P >` LINE CONTROL `< BR >` COMMANDS

显示为:

THIS IS

AN EXAMPLE

OF THE

LINE CONTROL

COMMANDS

`< P >` 标记的一般使用方法为: `< P ALIGN = # > ... </P >`。

属性 `ALIGN` 用于控制段落的排列方式, 它有 3 个属性值可选: `center` (居中)、`left` (左对齐) 和 `right` (右对齐)。

实际应用中, 有时省略了段落结束标记 `</P >`, 因为一个段落

的结束意味着下一个段落的开始，这样省略结束标记而由 <P> 标记取代。注意最后一个段落需要使用 </P> 标记，以保证段落终止和结构的完整性。

另一个有用的分隔文本的标记是 <HR> (水平控制)，将 <HR> 置于文档的任何地方，将使得 <HR> 上面的文本和下面的文本被一条屏幕上的从左至右的线分隔开来，可应用 <HR> 来分隔页面里的章节，修饰页面。它是一个单边标记，即无相应的终结标记 </HR>。<HR> 标记可带多个属性来对其长、宽、高以及对齐方式进行修饰，其格式为：

```
<HR [ SIZE = n1 ] [ WIDTH = n2 ] [ ALIGN = # ] [ NOSHADE ]  
[ COLOR = #rrggbb ] >
```

表 1.2 为 <HR> 标记的属性列表。

表 1.2 <HR> 标记的属性列表

属 性	意 义
SIZE	描述分割线的粗细（即水平线的高度），如：<HR SIZE = 6>
WIDTH	描述分割线的宽度（即水平线的长度），如：<HR WIDTH = 80>，表示画屏幕宽的 80% 长的一根水平线
ALIGN	描述分割线对齐的方式（左、右、中），如 <HR ALIGN = left>
NOSHADE	以没有阴影的方式显示分割线，即表示画一实心线段
COLOR	设置水平线的颜色，它的属性值为 rrggbb，是十六进制 RGB 数码，或者是下列预定义色彩：Black, Olive, Teal, Red, Blue, Maroon, Navy, Gray, Lime, Fuchsia, White, Green, Purple, Silver, Yellow, Aqua。该属性可用 COLOR = #rrggbb 形式表示，也可用 COLOR = "rrggbb" 形式

文字的分区 (Division) 显示采用 <DIV> 标记, 格式为:

```
< DIV ALIGN = left/center/right > ... </DIV>
```

left、center、right 分别表示将...对应的文字内容靠左、中、右显示。

整块引用采用 <BLOCKQUOTE>...</BLOCKQUOTE>, 功能类似 <PRE>...</PRE>。

此外, 可以用 <MULTICOL> 标记表示分栏, 格式为:

```
< MULTICOL [ COLS = n1 ] [ GUTTER = n2 ] [ WIDTH = n3 ] > ... </MULTICOL>
```

其中 n1 为列的数目, n2 为列间的空白, n3 为列的宽度。

1.2.1.3 列表结构

如果不使用 <PRE> 和 </PRE>, 就不能在文本中放入连续的空格符, 因为浏览器将删去多余的空格。然而, 列表结构能够将左边界向右移动。增强信息的条理性和层次性, 使浏览者一目了然, HTML 语言提供了 3 类列表结构。

(1) 无序列表

用 和 标记, 以圆点等符号为起始的列表结构。如○○○形式逐行列出项目, 每项一般要求以 标记开始。使用格式如下。

```
< UL TYPE = # >
  < LI > ...
  < LI > ...
  :
</UL>
```

这里 TYPE 是 标记的属性, # 为属性值。TYPE 属性的使用方法为:

< UL TYPE = disc >	表示以实心圆点符号为起始符号 (默认值)
< UL TYPE = circle >	表示以空心圆符号为起始符号 (实际

显示为一个空心方块)

`<UL TYPE = square >` 表示以实心方块符号为起始符号

`<LI>` 叫做定制列表元素, 使用格式为 `<LI TYPE = #>`, `# = disk, circle, square`。TYPE = disc 时以 * 标记排列, TYPE = circle 时以 o 标记排列, TYPE = square 时以 + 标记排列。也可采用 `<LI TYPE = #>` 格式, `# = A, a, I, i, 1` 分别指定用大写字母、小写字母、大写罗马数字、小写罗马数字、数字标记排列。

(2) 有序列表

用 `<OL>` 和 `</OL>` 标记, 以阿拉伯数字 1、2、3 … 形式逐行列出项目, 每项一般也要求以头标 `<LI>` 开始。使用格式为:

`<OL [TYPE = #] [START = ##] >`

`<LI> …`

`<LI> …`

`:`

`</OL>`

这里 TYPE 是 `<OL>` 标记的属性, # 为属性值。TYPE 属性的使用方法为:

`<OL TYPE = 1 >` 表示以阿拉伯数字 (如 1, 2, 3…) 为起始符号 (默认值)

`<OL TYPE = a >` 表示以小写字母 (如 a, b, c…) 为起始符号

`<OL TYPE = A >` 表示以大写字母 (如 A, B, C…) 为起始符号

`<OL TYPE = i >` 表示以小写罗马字母 (如 i, ii, iii…) 为起始符号

`<OL TYPE = I >` 表示以大写罗马字母 (如 I, II, III…) 为起始符号

START 属性用于指定项目符号的初始值, 例如 START = 3, 则对应的上面五种项目符号分别以 “3”、“c”、“C”、“iii”、“III” 开始。