

HTML

资源手册

[美] IAN S. GRAHAM 著

张文旭 李也白 译
鲁百年 刘红英



WILEY



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: <http://www.phei.co.cn>

HTML 资源手册

HTML SOURCEBOOK

[美] IAN S. GRAHAM 著

张文旭 李也白 译
鲁百年 刘红英



电子工业出版社

Publishing House of Electronics Industry

Copyright © 1995 Ian S. Graham

Published by John Wiley & Sons, Inc.

All rights reserved.

Authorized translation from English language edition published by John Wiley & Sons, Inc.

本书中文专有翻译出版权由美国 John Wiley & Sons 公司授予电子工业出版社。未经允许，不得以任何形式和手段复制或抄袭本书的内容。版权所有，侵权必究。

JS00067

书 名:HTML 资源手册

HTML SOURCEBOOK

著 者:[美]IANS. GRAHAM

译 者:张文旭 李也白 鲁百年 刘红英

责任编辑:秦 梅

特约编辑:李彦东

印 刷 者:北京科技大学印刷厂

装 订 者:三河赵华装订厂

出版发行:电子工业出版社出版、发行 URL:<http://WWW.phei.co.cn>

北京市海淀区万寿路 173 信箱 邮编 100036 发行部电话 68214070

经 销:各地新华书店经销

开 本:787×1092 毫米 1/16 印张:17.75 字数:454.4 千字

版 次:1997 年 4 月第 1 版 1997 年 4 月第 1 次印刷

印 数:5000 册

书 号:ISBN 7-5053-3814-5
TP·1634

定 价:29.00 元

著作权合同登记号:图书 01-96-1031

凡购买电子工业出版社的图书,如有缺页、倒页、脱页者,本社发行部负责调换

版权所有·翻印必究

目 录

引言.....	1
统一资源定位器.....	1
超文本传输协议.....	1
超文本置标语言.....	2
本书概要.....	2
致谢.....	4
献辞.....	5
第一章 超文本置标语言入门.....	7
超文本置标语言的基本概要.....	8
例 1：一个简单的 HTML 文档.....	8
HTML 单元.....	11
HEAD 和 TITLE 单元.....	11
BODY 单元.....	11
高亮度单元.....	12
段落.....	12
无序列表.....	13
横向标尺单元.....	14
例 1 小结.....	14
例 2：图象和超文本链接.....	15
文档范例.....	15
文档范例的显示.....	16
定位点.....	18
不完整的 URL.....	19
建立链接.....	21
超文本链接：好的、坏的和丑陋的.....	22
例 2 小结.....	25
例 3：主页.....	26
TITLE 和 HEADING.....	29
文本部分.....	30
组织.....	30
图标.....	31
统一资源定位器.....	32
例 3 小结.....	33
例 4：收集超文本文档.....	33
PRE 单元.....	36
组织.....	39
建立卷宗.....	39
例 4 小结.....	41
例 5：图象、电影和声音文件.....	41

链接大幅图象	41
辅助性应用程序	44
例 5 小结	44
例 6：表单	45
FORM 单元	45
表单的局限性	47
例 6 小结	48
参考文献	49
第二章 HTML 详解	错误！未定义书签。
HTML 介绍	51
HTML 文档允许的字符	52
特殊字符	52
HTML 文档中的注释	53
HTML 作为一个 MIME 类型	53
HTML 单元和置标标志符	54
字符大小写	55
空白单元	56
单元嵌套	56
未知的单元或属性	56
整体文本结构	56
超文本置标语言规范：单元详解	57
本节要点	57
HEAD 单元	59
BODY 单元	63
列表单元	76
字符有关的单元	89
字符高亮度单元	92
逻辑高亮度	92
物理高亮度	96
HTML+单元	97
参考文献	103
第三章 统一资源定位器 (URL)	105
URL 中允许的字符	105
不允许的字符	106
特殊字符	106
一个统一资源定位器的例子	107
1. 协议	107
2. 地址和端口号	107
3. 资源位置	107
不完整 URL	108
.....	108

URL 规范	109
FTP URL	110
GOPHER URL	111
HTTP URL	113
MAILTO URL	115
NEWS URL	115
TELNET/TN3270/RLOGIN URL	115
WAIS URL	116
FIEL URL	116
参考文献	116
第四章 HTTP 协议和通用网关界面	117
HTTP 协议	117
HTTP 客户机与服务器对话示例	118
用户鉴别, 数据加密和访问控制	128
HTTP 方法和消息头总结	130
HTTP 方法	130
HTTP 请求头	130
HTTP 响应头	131
通用网关界面	132
从客户机向服务器发送数据	132
从服务器发送数据到网关程序	132
从网关程序向服务器返回数据	133
POST 方法和标准输入	142
安全性	145
参考文献	146
第五章 HTML 和 CGI 工具	147
HTML 文档中的图象	147
X-BITMAP 图象	147
X-PIXELMAP	147
GIF 图象文件	147
缩小图象文件的大小: 颜色映射	148
减小图象尺寸: 改变图象比例	148
透明的 GIF 图象	149
活动图象	151
图标卷宗地址	155
客户机端的可执行程序	156
发送脚本到客户机	156
配置客户机	157
安全问题	158
服务器端文档包含	159
包含命令格式	160

服务器端包含的范例	163
HTML 实用程序	165
DTD2HTML	165
HTML 表格转换器	165
HYPERMAIL	166
MHONARC: MAIL TO HTML ARCHIVE	166
目录生成器	166
TREELINK	167
CGI 实用函数	167
CGI EMAIL 处理器	167
CGI 反馈表单	167
确定客户机软件	168
把 UNIX 手册转换成 HTML	168
处理查询和表单的软件包	169
数据库 CGI 网关程序	170
WAIS 网关程序	170
结构化查询语言数据库的网关程序	172
CGI 卷宗地址	175
数据库网关程序参考文献	175
第六章 HTML 编辑器和文档转换器	177
HTML 编辑器	177
简单的文本编辑器(MACINTOSH、PC、UNIX)	178
ALPHA(MACINTOSH)	178
BBEDIT HTML 扩展(MACINTOSH)	178
CU_HTML.DOT(PC-WORD FOR WINDOWS)	179
EMACS(UNIX,PC)	180
GT_HTML.DOT(PC-WORD FOR WINDOWS)	180
HOTMETAL(PC-WINDOWS、UNIX)	180
HTML ASSISTANT(PC-WINDOWS)	181
HTMLLED(PC-WINDOWS)	181
HTML.EDIT (MACINTOSH)	182
HTML EDITOR (MACINTOSH)	182
HTML FOR WORD 2.0(PC-WORD FOR WINDOWS)	182
HTMLTEXT	183
NEXTSTEP HTML-EDITOR(NEXT)	183
S.H.E. (MACINTOSH)	183
TKHTML (UNIX)	183
TKWWW (UNIX)	184
文档翻译器和文档转换器	184
ASC2HTML	185
CHARCONV	185
CYBERLEAF	186
FASTAG	186

FRAME2HTML.....	186
HLPDK.....	186
HYPERLATEX.....	187
LATEX2HTML.....	187
MIF2HTML.....	187
MIFTRAN	187
MM2HTML.....	188
MS2HTML.PL.....	188
PS2HTML	188
ROSETTAMAN.....	189
RTFTOHTML.....	189
RTFTOHTM.....	189
SCRIBE2HTML.....	189
STRIPHTML	190
TAGWRITE.....	190
TEX2RTF.....	190
TEXI2HTML	190
WPTOHTML.....	191
WP2X.....	191
WEBMAKER	191
HTML 检验工具	191
SGMLS.....	192
HTML 校验地址	193
链接校验工具	194
LINKCHECK.....	194
VERIFY_LINKS	194
第七章 WEB 浏览器和辅助应用程序	197
访问软件.....	197
PC 平台浏览器.....	197
MSDOS 浏览器: DOSLYNX	198
WINDOWS 和 OS/2 浏览器	198
辅助应用.....	201
MACINTOSH 平台浏览器.....	202
MACWEB	203
MACINTOSH 辅助应用	203
UNIX 和 VAX/VMS 平台浏览器.....	204
批处理模式浏览器(UNIX).....	204
CHIMERA (UNIX)	205
EMACS-W3 模式(UNIX、VMS、其它).....	205
LINEMODE 浏览器(UNIX、VMS).....	206
LYNX(UNIX、VMS).....	206
MIDASWWW (UNIX、VMS).....	206
MOAIC FOR X-WINDOWS (UNIX)	207
MOAIC(TUEV) 2.4.2	208

QUADRALAY GWHIS 浏览器(UNIX).....	208
RASHTY VMS 客户机(VMS).....	208
TKWWW(UNIX).....	208
VIOLAWWW(UNIX).....	209
UNIX 辅助软件包.....	209
NEXT 平台浏览器.....	210
CERN NEXT 浏览器.....	210
OMNIWEB.....	211
AMIGA 浏览器: AMOSAIC.....	211
诱人的前景.....	212
IBM OS/2 浏览器: WEB EXPLORER.....	212
MICROMIND SLIPKNOT.....	212
NETSCAPE 通信公司.....	212
第八章 HTTP 服务器和服务器实用程序.....	213
基本服务器问题.....	213
UNIX 服务器.....	213
VMS 服务器.....	214
WINDOWS NT 或 OS/2.....	214
WINDOWS 或 MACINTOSH.....	214
在防火墙后面?.....	214
服务器软件列表.....	215
CERN HTTP 服务器(UNIX、VMS).....	215
CL-HTTP(符号 LISP 机器).....	215
CMS HTTPD (VM/CMS).....	216
DECTHREAD HTTP 服务器(VAX/VMS).....	216
GN GOPHER/HTTP 服务器(UNIX).....	217
GWHIS HTTP 服务器(UNIX).....	217
HTTPS(WINDOWS NT).....	217
JUNGLE(UNIX).....	218
MACHTTP(MACINTOSH).....	218
NCSA HTTPD(UNIX).....	218
OS2HTTPD(OS/2).....	219
PLEXUS(UNIX).....	219
SERWEB(WINDOWS 3.1).....	220
WEB4HAM(WINDOWS 3.1).....	220
WINHTTPD: NCSA HTTPD FOR WINDOWS (WINDOWS 3.1).....	220
诱人的前景.....	220
BASIC WEBSERVER.....	221
MDMA: 多媒体访问的多线程监控程序.....	221
NETSCAPE NETSITE.....	221
服务器支持程序.....	221
GETSTATS.....	221
GWSTAT.....	222

WEBSTAT.....	222
WUSAGE.....	222
WWWSTAT.....	223
第九章 现实世界中的例子.....	225
电子出版服务器(E-PRINT).....	225
ONCOLINK.....	228
引言.....	229
ONCOLINK 的实现.....	229
ONCOLINK 的使用.....	231
太阳系的风景.....	235
背景.....	235
HTML 问题.....	237
总结.....	238
NETBOY□ □ 联机时代的选择.....	238
洛山矶预定处的 WORLD WIDE WEB 页.....	241
引言.....	241
为什么要在 WORLD WIDE WEB 上.....	241
设计 WWW 页面.....	242
对 SFR 的最后几点说明.....	248
附录 A ISO LATIN-1 字符集.....	249
URL 字符编码.....	249
HTML 字符和实体编码.....	249
附录 B 多义 INTERNET 邮件扩展(MIME).....	255
MIME 内容类型.....	255
服务器是如何确定内容类型的?.....	256
客户机是如何确定内容类型的?.....	256
参考文献.....	258
附录 C 使用 ARCHIE 寻找软件.....	259
标准 ARCHIE.....	259
MACINTOSH 的 ARCHIE.....	260
用 FTP 下载文件.....	261
病毒防护.....	263
附录 D 监听 TCP/IP 端口.....	265
词 汇 表.....	267

引言

完全可以说 World Wide Web 在 Internet 上掀起一场风暴, 它使 Internet 的怀疑者和支持者都同样目瞪口呆。然而在事后看来, 原因是一目了然的。不管是对 Internet 的信息用户还是对信息提供者来说, World Wide Web(WWW)模型都使他们对 Internet 的访问更为容易。它使得通过 Internet 发布信息变得那么容易, 搜寻起来也是那么简便。因此在不到三年的时间里, World Wide Web 工具迅速发展壮大, 成为 Internet 上最常用的工具是一点也不奇怪的。

好用的工具常常需要技巧和培训才能用好。利用超文本文档和 Internet 超文本服务器制作与发布信息所要使用的工具也是这样的。制作设计良好、实用和可靠的资源需要深入理解这些工具的工作方法以及如何才能用好这些工具。本书的目的就是帮助读者理解这些问题。假设读者已经熟悉 Internet 上传统的资源, 如 FTP、telnet、电子邮件、Gopher, 基本上还有以下三个方面需要考虑:

1. 统一资源定位器或 URL, 它是 WWW 中定位 Internet 资源的方案。
2. 超文本传输协议(HTTP)和 HTTP 客户机-服务器交互作用。HTTP 服务器是特殊设计的, 专门用于发布超文本文档, 读者必须了解它们是如何工作的, 以便充分发挥它们的优势。
3. 超文本置标语言或 HTML。World Wide Web 使用这一置标语言书写超文本文档, 而且它允许创建超文本链接、表单和单击式图象。书写好的 HTML 文档不仅涉及到技术问题(HTML 文档的合适结构), 还涉及到设计问题(确保把信息内容清晰地提供给读者)。本书旨在详细阐述以上三个问题, 并给读者提供开发高质量 World Wide Web 产品的工具。引言的以下各部分对这三个方面做一概括介绍, 解释它们的基本特征。然后是本书的纲要, 并对如何更好地探讨本书的内容和范例提出几条建议。

统一资源定位器

统一资源定位器或 URL, 它是个命名方案, 规定如何和到哪里去寻找 Internet 服务器(如 Gopher、FTP 或 WAIS 服务器)资源。例如: 访问匿名 FTP 服务器 ftp.bozo.net 上目录 /pub/web/browsers 下的文件 macweb.zip 的 URL 可以简单的写成:

`ftp://ftp.bozo.net/pub/web/browsers/macweb.zip`

WWW 超文本文档使用 URL 访问其它超文本资源。

超文本传输协议

超文本传输协议或 HTTP 是一个新的专为快速发布超文本文档而设计的 Internet 协议。和其它 Internet 工具如 FTP、WAIS 或 Gopher 一样, HTTP 也是一个客户机-服务器协议。在客户机-服务器模型中, 客户机程序在用户的机器上运行, 向运行在 Internet 其他机器上的

服务器程序发送消息请求服务。服务器向客户机回送一个消息响应请求。客户机和服务器都需要一个能很好理解的协议来交换消息。FTP、WAIS 和 Gopher 和其它的一些 Internet 客户机-服务器协议，World Wide Web 浏览器都能访问。然而 HTTP 协议是专为发布超文本文档而设计的，因此大部分通信都在 HTTP 服务器间进行。

在最简单的层次上，HTTP 服务器象匿名 FTP 服务器一样，在客户机提出请求时发送文件。然而 HTTP 服务器还支持其它一些重要的特性：

- 它不仅能为客户机返回文件，还能返回运行在服务器上的程序产生的信息。
- 能够收集来自客户机的数据，并把这些信息送往服务器上的其他程序进行进一步的处理。

这些特殊的服务器程序叫做网关程序，其作用是把 HTTP 服务器和其它局部资源如数据库链接起来。正象 FTP 服务器可以访问许多文件一样，HTTP 服务器也可以访问许多不同的网关程序；在这两种情况下，读者都可以通过 URL 指定想要的资源(文件或者程序)。

服务器和这些网关程序间的交互作用是由通用网关界面(CGI)规范来控制的。使用 CGI 规范，程序员可以很容易的书写简单的程序或脚本处理用户的查询、访问数据库、使图象响应鼠标的单击等等。

超文本置标语言

超文本置标语言或 HTML 是一种用于制作超文本文档的语言。正是这些文档在 World Wide Web 上发布，它们也正是读者在客户机上实际看到的东西。HTML 包括命令，也叫标志符，它把文本的标题、段落、列表、引文、强调等等标志出来。它还有其它一些标志符，它们分别用于在文档中插入图象，接受用户输入的表单，还有更重要的，用来包括超文本链接，把正在阅读的文档与其它文档或 Internet 资源如 WAIS 数据库和匿名 FTP 地址链接起来。正是这一点可以使用户单击一个高亮度的字符串就能够访问远在千里之外的计算机上的新文档、图象或电影文件。HTML 文档是如何指定这些文档的位置的呢？是 URL，它包含在 HTML 置标语言中，浏览器用它来寻找指定的资源。

URL 可以指向哪些资源？可以是其它的 HTML 文档、图象、声音文件、电影文件，甚至可以是数据库搜索工具。这些资源可以在用户自己的计算机上，也可以在 Internet 上其它任何地方。可以从 HTTP 服务器上访问，也可以从 FTP、Gopher、WAIS 或其它服务器上访问。URL 是一个非常灵活的方案，它和 HTML 相结合，形成一个惊人的强有力的软件包，生成一个超文本文档的互联网，使世界上所有超文本文档和 Internet 资源都相互链接在一起。事实上 World Wide Web 这个名字正是来自于这一资源互连的景象。

本书概要

本书介绍 HTML、URL、HTTP、CGI 接口，设计和制作通过 World Wide Web 发布的资源。本书首先从 HTML 语言开始。几乎所有的资源都是通过 HTML 文档来实现的，所以 HTML 展示就象人的脸一样。如果读者希望给观众留下一个深刻的印象，提供清晰简明的信息，关键是要掌握如何书写合适的 HTML，理解创建好的文档有关的各种设计问题。如果展示设计得不好，慢得令人沮丧或者难于理解，那么即使是世界上最好的 Internet 资源也难以为人们所接受。

HTML 也是一个明显的起点。读者可以先不去考虑 CGI 程序、HTTP 服务器和其它高级功能，就可以书写简单的 HTML 文档，并用 Mosaic、MacWeb、lynx、Cello 或 Netscape 等 WWW 浏览器观察。可以方便地在文档中加入指向世界各地的服务器资源的 URL，并熟悉系统是如何工作的。浏览器理解 HTML 超文本标志和它们包含的 URL，浏览器中内含的软件可以使用适当的协议与 Internet 服务器对话。这样仅仅创建少数几页 HTML 就可以完成一大堆工作。

第一章介绍 HTML 和制作 HTML 文档有关的设计问题。这一章是非技术性的，简要介绍 HTML 和文档设计过程中的几个重要方面。HTML 语言的详细内容和更为复杂的客户机-服务器问题留到第二章和第三章讨论。

设计问题对于开发好的 World Wide Web 展示至关重要。HTML 文档与文本文档或传统的超文本展示不同，它们受到浏览器能力的不同和 Internet 传输文档的速度的限制。第一章讨论这些问题在实际中的影响，并给出几条准则以避免 HTML 的主要著作错误。在大多数情况下，书中都使用正在讨论的重要问题作为例子以便读者能够迅速掌握其要点。

在第一章和本书各处都一直在强调，重要的是在创建 HTML 文档时要使用正确的 HTML 置标结构。尽管 HTML 是一个比较简单易懂的语言，但还是有许多重要的准则指定标志符的位置。只有确保文档遵循这些准则，才能保证它们可以在客户使用的不同浏览器上正确显示。制作的文档在一个浏览器上看起来非常好，但在其它的浏览器上却很难看甚至根本没法看是常有的事。

第一章包括了构造正确的 HTML 文档的几个基本准则，而第二章和其中的参考文献则是改正 HTML 的详细指南。特别需要指出的是，第二章详细阐述了 HTML 语言和其中各种允许嵌套的 HTML 置标指令。它还讲述了构造 URL 的语法和准则。

第三章讲述统一资源定位器(URL)的结构和语法。它是 HTML 文档中用以指明超文本链接的目标的寻址方案。

当然，HTML 仅仅是个开头。要真正发挥系统的优势，必须要掌握 WWW 客户机浏览器和 HTTP 服务器间的交互作用，还要能够编写服务器上的网关程序发挥这种交互作用的优势。第四章深入研究 WWW 客户机和 HTTP 服务器间的交互作用，讲述用于编写和 HTTP 服务器打交道

的服务器程序的通用网关界面(CGI)规范。第四章中有简单的例子说明 HTTP 协议和 CGI 接口,在参考文献中包含的 Internet 地址上可以找到交互式的教学文档。第二、三、四章是本书的技术核心,书写 HTML 文档和服务 CGI 程序时应当用它们作为参考资料。

不用说,从处理单击式图象的 `imagemap` 程序到复杂的数据库前端软件包, Internet 上已经有许多实用的 CGI 程序。所有这些实用 CGI 工具都放在第五章讨论,在开发和组织 HTML 文档过程中它们作为辅助工具也很有用。例如:有收集电子邮件信件转换成超文本档案的工具,也有为大量有关的 HTML 文件建立超文本目录的工具。第五章的第二部分讨论这些工具,指出它们所在的地址。它们中大部分都可以通过 Internet 从匿名 FTP 地址或 HTTP 服务器上得到。这些程序和现有的辅助性文档的地址都用 URL 给出。

因为 HTML 置标标志符是复杂的文本字符串,而且必须包含到文本文档中,所以制作 HTML 文档是一件非常烦人的事。HTML 代码的录入很费时间,而且经常是错误百出。所有这些标志符的录入都很容易出错。正因为如此,许多个人、团体和公司都正在开发 HTML 编辑器,以利于文档制作。第六章简要介绍可用于 PC、Macintosh 和 UNIX 平台的各种 HTML 编辑器,并说明如何得到它们。

在有些情况下,需要的可能不是一个编辑器,而是一个从一种格式到另一种格式的文档转换器,如从 Microsoft 的 Rich Text 格式、FrameMaker 的 MIF 格式、LaTeX 格式等转换成 HTML。第六章归纳了几个这样的软件包和得到它们的方法。最后是几个用于校验 HTML 文档与语言规范的一致性和检查文档内超文本链接的正确性的实用工具。这些工具列在第六章的末尾。

第七章简要回顾用于探索 WWW 和观看 HTML 文档的几个不同的浏览器。除自己经常使用的浏览器外,读者也应当对文档在其它浏览器上的显示效果感兴趣。如果单为某一个浏览器设计,就很有可能剪裁 HTML 以适应该特定程序的特殊性,这样做是非常危险的。因为这样就会导致 HTML 文档在自己的浏览器上看起来很好,但在其它的浏览器上却非常难看。同时使用另外一两个浏览器以避免这种错误是明智之举。

第八章讨论建立 HTTP 服务器的有关问题,并简要回顾一下常用的服务器软件包。这些服务器软件和辅助性文档的地址仍由 URL 给出,以供参考。

第九章是本书的最后一章,给出几个 WWW 地址的例子,其中包含令人感兴趣的和设计良好的展示。这些地址范围甚广,包括商业、娱乐、科学研究和教育。我请这些地址的创作者讲述了它们的制作过程,这样可以使读者对有关的问题有一个感性认识。鼓励读者访问这些地址,看看它们是如何工作的。保证会给读者留下深刻印象。

最后,走出去!去浏览!写本书把自己对好的或坏的设计的想法说出来固然很好,但作为 HTML 文档的写作者,更应当走出去通过观察和感受来体会它的效果。本书仅仅是一个框架,因此要想理解千百万人已经做的大量工作,只有走出去,自己看!

致谢

从 Ziegler 和 Terry Jones 那里我也受益匪浅，他们在图象文件转换和计算机网络的几个问题上给了我很大的帮助。

最重要的我要感谢我的妻子 Ann Dean，感谢她的编辑技术和她过去几个月中对我坚定不移的支持和耐心。没有她，也就不会有这本书。

献辞

谨以此书献给 Ann。

第一章 超文本置标语言入门

超文本置标语言或 HTML，用于指明文本文档的逻辑结构，并具有超文本链接和用户交互作用等重要扩充功能。HTML 并不是一种象 Word 或 WordPerfect 等那样的所见即所得 (WYSIWYG) 的字处理语言。相反，HTML 要求把文档分成不同的逻辑单元如题目、段落、列表，而把对这些标志单元的解释留给显示文档的浏览器。这个模型使系统具有巨大的灵活性，而且允许不同能力的浏览器观看同一 HTML 文档。事实上，从最高档的 UNIX 图形计算机到 VT-100 或老式的 8086 DOS 机的纯文本终端，都有各自的浏览器。例如：观看同一文档时，图形 UNIX 浏览器可能把主要的标题用大号字，也许是斜体或粗体(因为图形显示器可以显示优美的字型)显示。而 VT-100 浏览器可能仅仅是用它唯一可用的字体把题目居中显示。两种展示看起来不同，但它们都能把 HTML 标志符建立的逻辑结构再现出来。

HTML 文档看起来到底是什么样子？图 1.1 是个简单的例子。它看起来就象一个纯文本文档一样。事实上它本来也正是这样的。因此 HTML 文档可以用 Windows PC 上的 NotePad，Macintosh 上的 TeachTex 或 UNIX 工作站上的 vi 等简单的文本编辑器书写。建立 HTML 文档并不需要特别的文字处理器或高档的 HTML 编辑器。

HTML 文档和简单的文本文档的不同在于 HTML 置标标志符，就是那些用小于号和大于号(<……>)括起来的文本，也就是这些指令把文档中各部分的意义告诉浏览器。例如：标志符<H1>表示一级标题的开始，而</H1>标志符则表示一级标题的结束。因此字符串

```
<h1> This is the Heading </H1>
```

表明字符串” This is the Heading ” 是一个一级标题(标题总共有六级，从 H1 到 H6)。注意标志符中的斜杠表示结束标志符。标志符中名字的大小写是没有关系的，所以<h1>和<H1>，和是一样的。推荐用大写字母开头，以突出标志符。

HTML 文档有不同的单元构成。字符串

```
<h1> This is the Heading </H1>
```

就是一个 H1 单元，由 H1 起始标志符，封装的文本和 H1 结束标志符组成。H1 单元也常称为标题封装盒。

有些单元叫做空白单元，意即它们不对文本块发生作用，也不需要结束标志符。图 1.1 中的 BR 单元就是一个例子。标志符
在它所在的位置强制产生一个断行。断行不包含其它任何东西，因此也不影响任何封装的文本。

有时单元具有属性，它很象变量，通常也可以赋值，用以定义单元的特性。IMG 单元是例子，HTML 文档用它包含图象。IMG 单元可能以如下形式出现：

```
<IMG SRC= “ filename.gif ” >
```

这里 SRC 是 IMG 单元的一个属性，给出包含在文档中的图象文件的名字(实际上是一个指向图象文件的统一资源定位器(URL))。属性名字象单元名字一样与大小写无关。因此