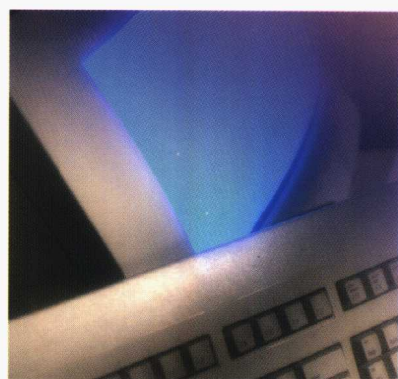
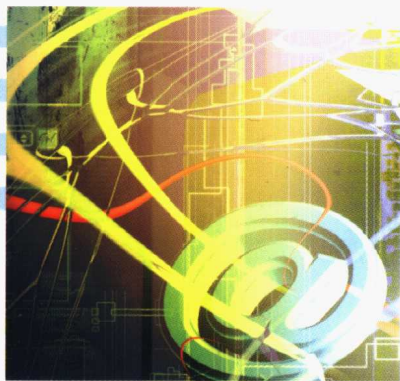
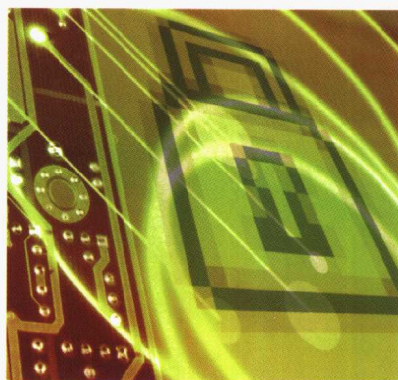
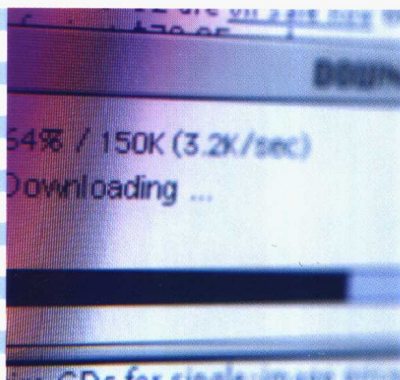




21 世纪高等学校应用型教材

Web 程序设计 教程

□ 主编 张念鲁 刘红屏



高等教育出版社
Higher Education Press

21 世纪高等学校应用型教材

Web 程序设计教程

主 编 张念鲁 刘红屏
副主编 马晓晨 赵传惠
王世辉 秦忠玉



高等教育出版社

内容提要

本书主要介绍 Web 程序设计的基本方法和技巧,全书共 9 章,主要内容包括:Web 程序设计集成开发环境、Web 脚本语言 VBScript、Web 脚本语言 JavaScript、ASP 程序设计概述、ASP 内置对象、ActiveX 程序设计、Web 数据库访问技术、ASP 程序调试与错误处理和 Web 应用程序设计实例。

书中配有大量经典实用的例题和习题供读者练习使用,可以帮助读者在较短的时间内掌握 Web 编程的基本方法和技术。特别是书中的程序范例“远程教育系统”,对读者开发 Web 实际应用程序起到指导作用。

本书结构清晰,内容由浅入深。书中程序均来自编者的开发实践。本书既可作为成人或高职高专学校相关专业的网络编程教材,也可作为网络编程技术人员和爱好者的参考书。本书配套电子课件及书中案例的源代码可从高等教育出版社网站下载,网址为 <http://www.hep.edu.cn> 或 <http://cs.hep.com.cn>。

图书在版编目 (CIP) 数据

Web 程序设计教程 / 张念鲁, 刘红屏主编. —北京:
高等教育出版社, 2004. 1

ISBN 7-04-013371-7

I. W... II. ①张...②刘... III. 计算机网络—程
序设计—教材 IV. TP393.09

中国版本图书馆 CIP 数据核字 (2003) 第 121652 号

出版发行	高等教育出版社	购书热线	010-64054588
社 址	北京市西城区德外大街 4 号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-82028899		http://www.hep.com.cn

经 销 新华书店北京发行所
印 刷 北京外文印刷厂

开 本	787×1092 1/16	版 次	2004 年 1 月第 1 版
印 张	21.25	印 次	2004 年 1 月第 1 次印刷
字 数	520 000	定 价	27.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

前 言

WWW(World Wide Web)即万维网,又称 Web,它是在 Internet 上运行的覆盖全球的多媒体信息系统。人们利用 Web 工作、学习和生活,Web 已经成为人们生活的重要组成部分。Web 站点数量以惊人的速度增长,网站建设的技术也在高速发展,越来越多的人在学习使用 Web,目前,大学的计算机及相关专业大多开设了 Web 编程技术课程,但相应的教材却很匮乏。现有的 Web 编程技术相关教材多为专业技术人员编写。本书专门为大学的计算机相关专业的学生而编写。

本书全面介绍了创建和维护 Web 站点所需要的编程工具和技术,重点介绍如何使用 ASP 来构建动态交互的网站,并以案例驱动的形式详细介绍了 Web 编程技术。全书共 9 章。第 1 章 Web 程序设计的基础知识,主要内容包括:Web 概述、Internet 概述、HTML、CSS 和 XML。第 2 章 Web 脚本语言 VBScript,主要内容包括 VBScript 编程基础、VBScript 控制流程结构、VBScript 的函数、过程、对象等。第 3 章 Web 脚本语言 JavaScript,主要介绍了 JavaScript 编程基础、JavaScript 控制流程结构、JavaScript 的函数、过程、对象等。第 4 章 ASP 程序设计概述,介绍几种常用 Web 动态服务器技术 ASP、JSP、PHP,重点介绍 ASP 程序设计的基础。第 5 章 ASP 应用程序设计,结合案例介绍了 ASP 内置对象,给出“远程教育系统”的总体设计。第 6 章 ActiveX 技术,详细介绍了 ASP 中常见 ActiveX 组件和 ASP 组件的创建等。第 7 章 Web 数据库访问技术,介绍了 ODBC、ADO、SQL 和 ASP 数据库访问技术以及“远程教育系统”的数据库设计。第 8 章 ASP 程序调试与错误处理,介绍 ASP 的错误类型、错误处理方法、ASP 的程序漏洞和解决办法,并提出了 ASP 的安全建议。第 9 章 Web 应用程序设计实例,详细介绍了搜索引擎、新闻发布、聊天室和 BBS 的程序设计。

读者通过学习本书可以掌握基本的网页编程技术;学会使用 Web 脚本语言编写 Web 应用程序功能模块;掌握 Web 数据库访问技术和方法;熟悉利用 ASP 开发 Web 应用程序的全过程。

本书要求读者具备一定的面向对象程序设计、计算机网络和数据库的基本知识,以及网站建设的基本常识。

本书由高校计算机专业具有丰富 Web 教学和开发经验的教师编写,案例源自实际应用系统,所有的代码都通过调试。全书整体框架和编写思路由张念鲁设计并最后统稿。参加本书编写和程序调试的有:张念鲁、刘红屏、马晓晨、赵传惠、王世辉、秦忠玉、宋文津、陈宏、乔显亮、王晓娟、翟乃强、冯文仙、杜海平等。在本书的编写过程中,朱彦之女士和张云开先生给予了大力的帮助和支持,在此表示由衷的感谢。

由于编者的水平有限,恳请读者对本书提出宝贵意见和建议。编者 E-mail:nianlu369@sohu.com。

编 者

2003 年 11 月

目 录

第 1 章 Web 程序设计基础	(1)	2.2.2 常量和变量	(55)
1.1 Web 概述	(1)	2.2.3 运算符和表达式	(58)
1.1.1 Web 的工作方式	(1)	2.2.4 数组	(62)
1.1.2 Web 页面的特点	(2)	2.3 VBScript 程序流程控制	(66)
1.1.3 Web 页面的功能	(2)	2.3.1 条件控制语句	(66)
1.2 Internet 概述	(3)	2.3.2 循环控制语句	(71)
1.2.1 TCP/IP 协议	(3)	2.4 过程	(76)
1.2.2 HTTP 协议	(4)	2.4.1 Sub 过程	(76)
1.2.3 Telnet 协议	(5)	2.4.2 Function 过程	(80)
1.2.4 FTP 协议	(6)	2.5 常用内部函数	(81)
1.2.5 IP 地址	(6)	2.5.1 输入/输出函数	(81)
1.2.6 统一资源定位器 URL	(7)	2.5.2 字符串函数	(83)
1.3 超文本标记语言 HTML	(7)	2.5.3 日期函数	(92)
1.3.1 HTML 的基本语法	(9)	2.5.4 转换函数	(93)
1.3.2 HTML 文档的基本结构	(16)	2.5.5 VBScript 中的几个数学函数	(94)
1.3.3 编写简单的 HTML 文件	(17)	2.6 对象	(96)
1.4 样式表 CSS	(18)	2.6.1 文档对象模型(DOM)	(96)
1.4.1 样式表 CSS 的基本语法	(19)	2.6.2 Windows 对象	(97)
1.4.2 样式表的继承和作用优先级	(24)	2.6.3 Document 对象	(99)
1.4.3 CSS 的属性	(25)	2.6.4 History 对象	(101)
1.4.4 应用 CSS 的一个实例	(30)	2.6.5 Navigator 对象	(102)
1.5 可扩展标记语言 XML	(33)	2.6.6 Form 对象	(103)
1.5.1 XML 文档的组成	(33)	2.6.7 文本框对象	(105)
1.5.2 创建 XML 文档的基本规则	(34)	2.6.8 按钮对象	(109)
1.5.3 将格式正确的 XML 文档 转换为有效文档	(35)	2.6.9 单选框对象和复选框对象	(111)
1.5.4 定义和使用实体	(36)	2.6.10 选择列表对象	(115)
1.5.5 XML 文档的显示	(38)	2.6.11 Location 对象	(117)
1.5.6 XML 应用实例	(39)	2.7 数据有效性验证实例	(119)
习题一	(43)	2.8 ActiveX 控件	(121)
第 2 章 Web 脚本语言 VBScript	(44)	习题二	(122)
2.1 VBScript 语言概述	(44)	第 3 章 Web 脚本语言 JavaScript	(125)
2.1.1 VBScript 语言的特点	(44)	3.1 JavaScript 语言概述	(125)
2.1.2 VBScript 程序的编辑和调试	(45)	3.1.1 JavaScript 语言的特点	(125)
2.2 VBScript 编程基础	(53)	3.1.2 JavaScript 程序的编辑和调试	(127)
2.2.1 数据类型	(53)	3.2 JavaScript 编程基础	(128)
		3.2.1 数据类型	(128)

3.2.2 常量	(130)	6.2.4 CDO 组件	(216)
3.2.3 运算符和表达式	(131)	6.2.5 File Access 组件	(218)
3.3 JavaScript 中的对象	(133)	6.3 用 VB 创建 ASP 组件	(219)
3.4 JavaScript 程序流程控制	(139)	6.4 远程教育系统说明——学习 支持系统	(225)
3.4.1 条件控制语句	(140)	习题六	(230)
3.4.2 循环控制语句	(141)	第 7 章 Web 数据库访问技术	(231)
3.5 函数	(148)	7.1 Web 数据库访问技术简介	(231)
3.6 事件和事件过程	(151)	7.2 ODBC	(235)
3.6.1 文档对象模型用于脚本编写	(151)	7.2.1 ODBC 概念	(235)
3.6.2 数据有效性验证实例	(164)	7.2.2 配置数据源	(236)
习题三	(164)	7.3 ADO	(237)
第 4 章 ASP 程序设计概述	(167)	7.3.1 ADO 数据访问技术概述	(237)
4.1 动态服务器技术的产生背景	(167)	7.3.2 ADO 的对象类和对象模型	(239)
4.2 目前主要的动态服务器技术	(168)	7.4 结构化查询语言 SQL	(241)
4.2.1 ASP 概述	(168)	7.4.1 SQL 语言概述	(241)
4.2.2 PHP 概述	(168)	7.4.2 SQL 数据操作基础	(243)
4.2.3 JSP 概述	(169)	7.4.3 SQL 高级应用	(248)
4.2.4 各种动态网页技术的比较	(170)	7.5 使用 ASP 访问数据库	(251)
4.3 ASP 模型	(170)	7.5.1 ASP 与数据库的连接	(251)
4.4 ASP 的运行环境与设置	(171)	7.5.2 检索数据库内容	(252)
4.5 ASP 的主要内容	(174)	7.5.3 分页技术	(255)
4.6 程序范例	(178)	7.5.4 向数据库中添加新数据	(257)
4.6.1 范例介绍	(178)	7.5.5 修改数据库中已存在的数据	(259)
4.6.2 代码讲解	(179)	7.5.6 删除数据库中的数据	(260)
习题四	(180)	7.5.7 调用存储过程	(261)
第 5 章 ASP 内置对象	(181)	7.6 远程教育系统说明——数据库设计与 管理模块	(263)
5.1 Request 对象和 Response 对象	(181)	7.6.1 数据库设计	(263)
5.1.1 Request 对象	(181)	7.6.2 教师功能模块	(270)
5.1.2 Response 对象	(187)	7.6.3 管理员功能模块	(277)
5.2 Application 对象和 Session 对象	(189)	习题七	(284)
5.2.1 Application 对象	(189)	第 8 章 ASP 程序调试与错误处理	(286)
5.2.2 Session 对象	(194)	8.1 ASP 中的主要错误类型	(286)
5.3 Server 对象	(196)	8.2 防止错误的方法和良好的 编程习惯	(287)
5.4 远程教育系统说明——系统 总体说明	(199)	8.3 ASP 程序错误处理方法	(289)
习题五	(205)	8.3.1 错误处理	(289)
第 6 章 ActiveX 程序设计	(206)	8.3.2 程序调试——发现及处理 错误	(291)
6.1 ActiveX 组件的创建和使用方法	(207)	8.4 常见的 ASP 漏洞及解决方法	(294)
6.2 ASP 中常用的 ActiveX 组件	(209)	8.4.1 ASP 的安全优点	(294)
6.2.1 AD Rotator 组件	(209)		
6.2.2 Browser Capabilities 组件	(211)		
6.2.3 Counters 组件	(213)		

8.4.2 ASP 漏洞分析和解决方法	(294)	9.2 新闻发布系统程序设计	(313)
8.5 ASP 安全建议	(302)	9.3 聊天室程序设计	(322)
习题八	(310)	9.4 BBS 程序设计	(325)
第 9 章 Web 应用程序设计实例	(311)	习题九	(330)
9.1 搜索引擎程序设计	(311)	参考文献	(331)

第1章

Web 程序设计基础

本章导读

本章主要介绍 Web 程序设计所必需的基础知识,其中包括 Web 的基本概念和工作方式及特点、Internet 网络协议、IP 地址、与 Web 有关的协议和统一资源定位器 URL、超文本标记语言 HTML、样式表 CSS、可扩展标记语言 XML 的基本语法和规则,并给出一些简单的应用实例,为 Web 程序设计奠定基础。

1.1 Web 概述

WWW(World Wide Web)即万维网,又称 Web,它是在 Internet 上运行的覆盖全球的多媒体信息系统。它是由欧洲粒子物理实验室(CERN)研制的一个基于超文本(Hypertext)方式的信息检索服务工具。它提供了在 Internet 上的一种十分有效的浏览、检索及查询信息的方式,这些信息有多种类型,包括文字、图形、图像、声音和动画等。WWW 提供了内容丰富且又随时更新的信息资源,这些信息资源遍布在世界各地的 Web 服务器上。另外,WWW 还可提供“传统的”Internet 服务:Telnet、FTP、Gopher 和 Usenet News(Internet 的电子公告板服务)。

1.1.1 Web 的工作方式

Web 的信息源保存在 Web 站点中,用户通过 Web 浏览器(如 Microsoft 的 Internet Explorer 5.0)来访问。因此,Web 是一种基于客户机/服务器(Client/Server,C/S)的体系结构。用户使用浏览器从网上查阅 Web 信息,把需要的信息从网上下载到本机。由于信息的分布点不同,用户需求信息的不同,表现在 Web 上的是一种链接地址的不断变化。

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(Hyper Text Markup Language,超文本标记语言)写成的文档,包括嵌入在 HTML 文档中的 GIF 和 JPEG 格式的图像。此外,浏览器还可以根据用户的需要配置某些辅助应用程序,用来处理嵌入在 HTML 文档中的声音、视频等外部多媒体信息。人们通常将 Web 浏览器中显示的 HTML 文本称为 Web 页面(Page)。

在 Web 系统中,Web 服务器向浏览器提供服务的工作方式如下:

(1) 用户启动客户机(即本机)的浏览器程序,并在浏览器中指定一个 URL(Uniform Resource Locator,统一资源定位器),它是浏览器用来访问 Internet 信息的地址,即通常所说的网址。它准确地描述了信息所在的地址,浏览器便向该 URL 所指向的 Web 服务器发出请求。

(2) Web 服务器接到浏览器的请求后,把 URL 转换成页面所在服务器上的文件路径名。

(3) 若 URL 指向的是普通的 HTML 文档,Web 服务器直接送给浏览器,浏览器负责将 HTML 格式文档解释转换成用户能接受的文本格式。HTML 文档中可能包含有用 Java、JavaScript、ActiveX 或 VBScript 等编写的小应用程序,服务器也将它们随 HTML 一起传到浏览器,在浏览器所在的机器上执行。

1.1.2 Web 页面的特点

一个好的 Web 页面具有如下特点:

(1) Web 页面的可读性强。Web 是一种超文本信息系统,它不同于印刷品。印刷品有触感和随身携带特点,是一种固定的、线性的,它不能从一个位置迅速跳转到另一个位置,从一个主题迅速跳转到另一个相关的主题。而 Web 页面触感弱,主要靠键盘、鼠标、显示器将读者和页面内容连接起来。但由于它的超文本特性,用户通过点击相关的关键词的超链接,可以迅速实现在线帮助,提高阅读能力。

(2) Web 页面有导航特性。Web 页面可以将图形、音频、视频信息集合于一体。这完全是靠它的导航特性。一个页面具有同时显示图形和其他超媒体的性能。网站与网站之间、页面与页面之间的超链接,使用户浏览方式多样化、信息下载迅速快捷、访问距离缩短。这是 Web 页面的独特之处。

(3) Web 页面与系统平台无关。不论用户使用的操作系统平台是什么,均可通过 Internet 访问 Web,只要大家共同遵守 Internet 的协议即可。

(4) Web 页面信息的分布特性。对一个网站,不可能将大量的图形和音、视频信息放在一起。各站点数据库信息的交互驱动,使得物理位置不同的信息逻辑上一体化。对用户来说信息是一体的,只是通过超链接进行不同站点的访问而已。

(5) Web 页面的动态交互性。随着诸如 ASP 和 Java 等动态技术的发展,早期的静态 Web 页面已发展到现在的具有动态性、交互性的网页。

(6) Web 页面的新闻性。Web 站点的信息是动态的、经常更新的。信息的提供者可以经常对站点上的信息进行更新,用户可以得到最新的信息。

1.1.3 Web 页面的功能

Web 页面可实现的功能主要表现在以下三个方面。

1. 主页的功能

网站主页用于组织和表现内容,它提供到站点主要部分的链接,以此来指引用户。主页的合理设计非常重要,它反映客户的产品或服务的特点,是网站最重要的核心页面,是网站的灵魂所在。所以设计时要注意主页内容和样式的最佳平衡。一个主页上部有该站名称、标志图片。名

称下一行有横向或左侧纵向的动态折叠菜单,而每一个动态折叠菜单下有多个子菜单,它们之间都有超链接,浏览者可根据需要浏览相关内容。

2. 超链接功能

通过超链接功能,清晰的导航机制可迅速引导用户进入不同的网站或页面。

网页的基本要素有文字、图形和链接标识。文字有标题、字号和字型的不同。图形有菜单按钮和背景图形。而链接标识是 Web 页面非常重要的功能,也是基本元素之一。可以实现通过链接从一网页转到另一网页,从一网站转到另一网站。

3. 页面的交互性

交互是网站响应用户动作和选择的方式。交互式的网站会吸引更多用户的参与、允许用户选择自己想看的内容,而不是静态地显示信息。交互类型有交互式导航、交互式多媒体、交互式广告、用户间的交互等。例如用户间的交互包括论坛、聊天、视频会议等。Web 页面上包括的公告栏,又叫新闻组或论坛,在这里,用户可以公开发布信息,而其他用户可以阅读这些信息并进行回复。在聊天室用户之间可以实时交换文本信息或声音信息。视频会议是一种使用日益频繁的交互式功能,虽然会产生图像和声音的延迟,随着带宽和速度的提高会日臻完善,它的最大特点就是节约时间和金钱。

1.2 Internet 概述

Internet 是在目前全球性电脑信息网络中覆盖面最大、信息资源最丰富的网络系统。Internet 被认为是未来全球信息高速公路的雏形。在 Internet 网络上的 E-mail 功能,可以接收和发送电子邮件。利用 Internet 网络上的 Telnet 和 FTP 功能,还能直接使用远程电脑主机的软件系统,以及丰富的信息资源。对于科学研究的课题、论文、图书馆的藏书和各种科学杂志等图像文字资料,都可以使用 WAIS、Archie、Veronica、Jughead 等工具,依据查询关键字,查询和检索到它们。利用 Gopher 交互式检索工具和 WWW 广域网超媒体信息获取工具以及 NetscapeNavigator、Internet Explorer、Mosaic、I-Comm、Lynx 等资源浏览工具,就可以更方便地获得和浏览所需信息。通过 Internet 网络上的 Usenet News Group,还可以看到世界各地新闻、全球气象预报、重大体育比赛的报道和各类广告。Internet 网络的 Talk、IRC 功能,则为我们提供了与生活在地球另一端的朋友随时进行交谈的机会。目前正在对 Internet 网络进行声音和图像同步传送的深入研究,这预示着在网上召开视频电子会议已是近在眼前的事情。借助于多媒体技术,Internet 网络用户可以收看和收听网络上世界各地的动画、图像资料和有声资料。

1.2.1 TCP/IP 协议

Internet 共同遵守的网络协议是 TCP/IP (Transmission Control Protocol/ Internet Protocol)协议。。它是一个工业标准的网络通信协议集,是为广域网设计的,其中最重要的是 TCP 协议和 IP 协议。因此,通常将这诸多协议简称为 TCP/IP 协议。

TCP/IP 协议把整个网络分为四个层次,自上而下分别是:应用层、传输层、网络层和物理链

路层。它们都是建立在硬件基础层之上。图 1-1 给出了 TCP/IP 参考模型与 OSI 参考模型的对照。

1. 应用层

应用层是 TCP/IP 参考模型的最高层,它向用户提供一些常用应用程序,如电子邮件服务、网页浏览服务等。应用层包括了所有的高层协议,并且不断有新的协议加入。该层主要有:网络终端协议 Telnet,用于实现互联网中的远程登录功能;文件传输协议 FTP,用于实现互联网中交互式文件传输功能;简单电子邮件协议 SMTP,实现互联网中电子邮件收发功能;网络文件系统 NFS,用于网络中不同主机间的文件系统共享;域名服务系统 DNS,用于实现网络设备域名到 IP 地址的映射服务;超文本传输协议 HTTP,用于在 Web 浏览器和服务端之间传输 Web 文档。Web 程序设计较注重的是 HTTP 协议、Telnet 协议和 FTP 协议等。

2. 传输层

传输层又称 TCP 层,是计算机通信体系结构中最关键的一层,又是最复杂的一层。主要功能是负责应用进程之间的端到端(End-To-End)通信服务(最流行的是源主机的应用进程和目标主机的应用进程之间的可靠数据传输服务),完成端到端的差错控制和流量控制,保证传输无差错,保持顺序,无丢失或无重复等。传输层定义了两种协议:传输控制协议 TCP 与用户数据报协议 UDP(User Datagram Protocol)。

3. 网络层

网络层又称互联网协议 IP 层,负责处理互联网中计算机之间的通信,向传输层提供统一的数据包。IP 协议不检查某个数据分组是否真的到达目的地,这称无确认传输;也不关心分组穿过互联网络的路径,甚至不确保各分组能否以其发送的顺序到达收方,这又称为无连接协议。但 IP 协议执行的功能是将传送的一个数据分组划分为多个较小的数据分组,以便能通过各种类型的设备和网络,将它们重新组装起来。

4. 物理链路层

物理链路层包括物理层和数据链路层,是主机与网络的实际连接层。物理层是最底层。实际计算机的网络连接通常实现为网络接口卡及插入主设备的主板。主要功能是利用物理传输介质为数据链路层提供物理连接。它是所有网络的基础。数据链路层是在物理层提供的位流传输服务的基础上,在通信的实体之间建立数据链路连接,传送以帧为单位的数据,采用差错控制、流量控制方法,使有差错的物理线路变成无差错的数据链路。

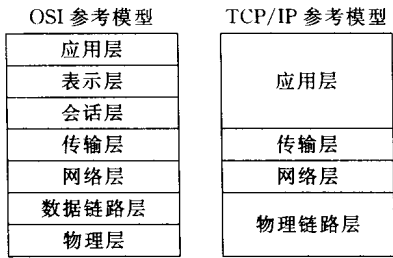


图 1-1 TCP/IP 参考模型与 OSI 参考模型的对照

1.2.2 HTTP 协议

HTTP(Hyper Text Transfer Protocol,超文本传输协议)是专门为 Web 设计的一种网络协议,属于 TCP/IP 参考模型中的应用层协议,位于 TCP/IP 协议的顶层,是在 Web 服务器和浏览器间的通信协议。HTTP 协议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 服务器和浏览器上广泛使用的版本是 HTTP/1.0[BFF96]或 HTTP/1.1[GMF99]。因此有时 Web 服务器被称为 HTTP 服务器。HTTP 需要可靠的双向字节流传输服务,是建立在 TCP 之上的应用层协议。HTTP 是基于客户和服务进行通信的基本模式,即请求回答模式。Web 浏览器和 Web 服务器之间的交互步骤如下:

- (1) 客户与服务器建立 TCP 链接;
- (2) 客户向服务器发出请求;
- (3) 服务器如果接受请求,就向客户发回响应信息,其中包括状态码和所要的文件(一般是 HTML 文档);
- (4) 客户与服务器关闭链接。

这里有必要简单介绍一下 Web 服务器和 Web 浏览器。

(1) Web 服务器

一个大学、研究机构公司或企业凡申请了网站,就有一个(或多个)Web 服务器。客户看到主页就基本能够了解该网站(Web 服务器)所提供的信息概貌。服务器上 Web 页的组织链接方式分层次式和线性式。一个 Web 文档一般组织成层次和线性的结合体,以方便客户对信息的阅读。Web 服务器上的信息不会包罗万象,大量的信息是世界范围内的所有 Web 服务器上 Web 页的集合,没有整体结构,将它们捆绑在一起靠的是超链接。这就是 Web 服务器需要一个 HTTP 服务器程序的原因。

Web 服务器在 TCP 端口 80 侦听来自 Web 浏览器的链接请求,浏览器根据用户的点击要求(如用户需要某个 Web 服务器上的特定文档时)向该服务器 80 端口发出 TCP 链接请求,链接建立之后,浏览器发送该文档的请求报文(其中主要是网址),服务器响应后抓取该文档返回给浏览器。

(2) Web 浏览器

Web 浏览器的基本功能是导航与浏览。导航是根据给定的超链接在 Web 中穿梭航行并从 Web 服务器获取 Web 页;浏览器负责显示 Web 页。导航是执行 HTTP 协议;浏览器是解释 HTML 语言。超链接所包含的指针用 URL 描述。用户浏览网页时常常看到一些加亮显示且带下划线的部分,鼠标一停留在此处,出现手形光标,一旦点击,浏览器马上去抓取该超链接所指向的 Web 页,显示给用户。

当然,HTTP 协议不仅仅是限于 Web 服务,还支持如 FTP 协议和 Telnet 协议及名字服务等。

1.2.3 Telnet 协议

Telnet 是 TCP/IP 的一个应用层协议,即 Internet 远程登录服务协议,是 Internet 上强有力的功能。要实现此功能,用户计算机上必需安装和运行名为 Telnet 的程序。使用该功能用户可以访问连在 Internet 上的任何一台 Telnet 服务器,用户主机成为该主机的终端,并且可以访问各种所需的信息。

- (1) 用户主机上装有包括 Telnet 应用层协议在内的 TCP/IP 协议族。
- (2) 要知道 IP 地址、域名、登录名、登录标识和口令,即所谓的权限。
- (3) 用户自己有账号,要完成拨号服务工作,使自己的调制解调器与远程主机建立通信关系。当然有的远程 Telnet 服务器对用户是开放的,不需用户有自己的账号、口令和用户名,照样可以登录,这就是为公众开放的 Telnet 远程服务。

1.2.4 FTP 协议

FTP(File Transfer Protocol,文件传输协议)是 TCP/IP 网络体系结构传输层的一种协议,用来提供计算机在 Internet 上传输二进制文件(程序和文档)的服务。它是一种上传/下载协议。通过 FTP 服务,允许 Internet 用户两地间的双向文件传输,实现真正意义上的全球资源共享。这些文件由文字的和非文字信息组成,包括:文档、程序、静态图像、照片、音频和视频信息。FTP 与 Telnet 的不同之处是:前者可以实现两地间的信息拷贝(在首先取得对方允许的情况下),而后者是只能在取得远程主机允许的情况下(如账号、口令、用户名登录)让用户主机以终端方式共享远程主机上的资源,而不能将远程主机上的文件资源复制到用户主机再传入用户自己的 PC 机上。

FTP 是一种文件传输协议,体现的是服务。提供服务的机器称为 FTP 服务器,又称 FTP 网站。这类网站有两类:一类是访问该服务器时必须拥有远程计算机上的合法账户和口令,且拥有相应的访问权限。另一类是匿名 FTP,即这类 FTP 服务器允许没有账户和口令的用户仍然可以从远程计算机上获取文件,其目的是向公众提供文件拷贝服务。在与这类“匿名”FTP 服务器建立连接时,只在“用户名”栏填入“anonymons”,在密码栏填入用户的 E-mail 地址作为口令即可完成注册。通过“匿名”FTP Internet,用户可免费下载程序、文件、数据和新软件。

1.2.5 IP 地址

IP 地址是互联网地址,每台连接于 Internet 上的主机要和其他连于 Internet 上的主机进行通信,必须要有一个惟一地址,这就是 IP 地址。这就像用街道地址来确定家庭确切物理位置一样,以确定主机在互联网上的位置。IP 地址有统一的格式,它是由协议 IP 规定的。

一个 IP 地址由 4 个字节组成,共 32 个二进制的。各字节间用“.”分隔,反映在十进制上就是 3 位数一组,共 4 组,如 191.168.1.50。

每一个 IP 地址都由两部分组成:网络 ID(NetWork ID)和主机 ID(Host ID)。网络 ID 代表一个网络,每个网络区段中的每台计算机都被赋予一个主机 ID。

IP 协议规定了 A~E 五类 IP 地址。其中 A、B、C 三类是基本的,它们由 IP 地址的高位来区分。格式如下:

	0	1		7	8		31
A 类	0		Net ID				Host ID
	0	1	2		15	16	31
B 类	1	0		Net ID			Host ID
	0	1	2		23	24	31
C 类	1	0		Net ID			Host ID

A 类 IP 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达 $2^{24} - 2 \approx 1678$ 万台。

B 类 IP 地址共有 $2^{14} = 16384$ 个,适用于中等规模的网络,每个网内主机数可达 $2^{16} - 2 = 65534$ 台。

C 类 IP 地址最多,有 $2^{22} \approx 419$ 万个,主要分配给小型网络使用,每个网内主机数最多为 $2^8 - 2 = 254$ 台。

1.2.6 统一资源定位器 URL

统一资源定位器 URL(Uniform Resource Locator)是用来确定用户所需信息的确切位置。包括信息资源的名称及地址,类似邮政地址或电话号码。

Web 浏览器为了进行浏览,首先要知道要访问的资源类型、资源地址、文档目录、文档名称。这就像指定一个人需说明他的国籍、省份、城市、街道、门牌号码一样,URL 用来指定 Internet 资源要说明它位于哪台计算机的哪个目录中。URL 通过定义资源位置的抽象标识来定位网络资源,格式如下:

访问方法://服务器域名[:端口]/目录/文件名

访问方法说明如何访问文档,即用什么协议或信息服务类型,可以是下列关键字:

HTTP:检索 Web 服务器上的文档(即超文本传输服务)。

FTP:检索匿名 FTP 服务器上的文档(即文件传输服务)。

GoPher:检索 GoPher 服务器上的文档(即 GoPher 服务)。

WAIS:检索 Wais 服务器上的文档(即提供检索数据库信息服务)。

NEWS:读取最新的 USENET 新闻(即提供网络新闻服务)。

TELNET:登录到某服务器(即远程登录服务)。

MAILTO:检索电子邮件地址。

冒号和“//”后部分指示一个网络主机的域名或 IP 地址,即文件所在的服务器。端口部分是可选项,一般情况下可省略。但在一些情况下指定端口是为了让操作系统用来辨认特定信息服务的软件端口,默认的是用于访问关联的端口,如 HTTP 默认的 TCP 端口是 80。

//目录/文件名是指文档在服务器中的目录。目录可包含子目录,各级目录之间用“/”分隔。下面举出一些 URL 应用例子:

http://www.cctv.com

telnet://odysseus.circe.com;70

ftp://ftp.w3.org/pud/www/doc

gopher://gopher.internet.com

news://comp.sys.novell

wais://quake.think.com/directory-of-servers

1.3 超文本标记语言 HTML

HTML 语言是一种简单、通用的标记语言,又称 Web 网页描述语言。它不同于 C/C++、Pascal 等程序设计语言。我们知道使用排版语言工具(如 WPS、Word)制作文档时,需要通过一

些命令按钮或菜单给文档加上一些控制码达到所需输出结果,如字体、字型、字号、底纹、段落设置等。而编制 HTML 文档时也需要加入一些类似的标记(Tag),用来说明段落、标题、字体、图形等。用 HTML 语言编制的文档为 HTML 文档,其文档扩展名通常是 htm 或 html。由 Web 浏览器来解释并显示在用户浏览器窗口中的页面是 Web 页面。带有标记的 HTML 文档是呆板的文档,而通过 Web 浏览器解释后呈现给用户的却是幅生动活泼、别开生面的画面。图 1-2 所示是一个用 HTML 语言编写的 HTML 文档,而图 1-3 所示是通过 Web 浏览器解释后呈现给用户的页面。

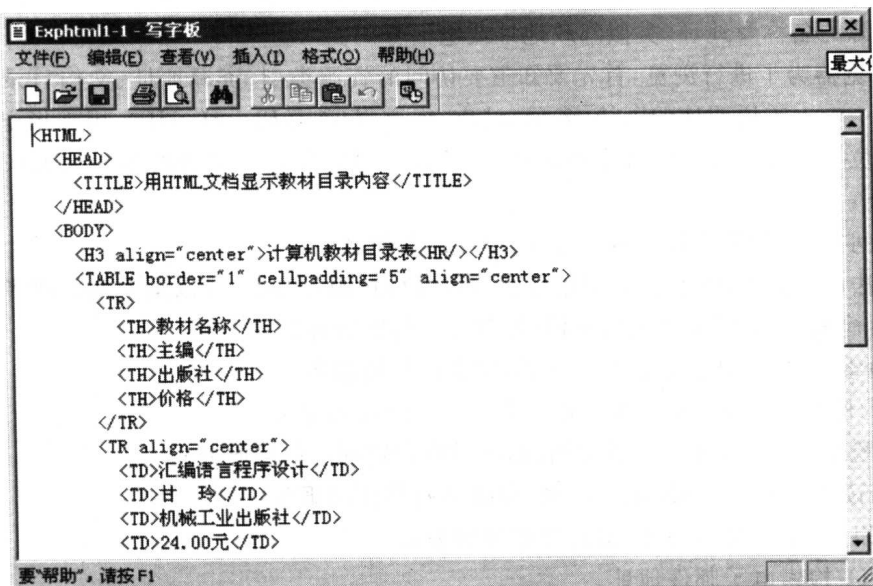


图 1-2 用 HTML 语言编写的 HTML 文档

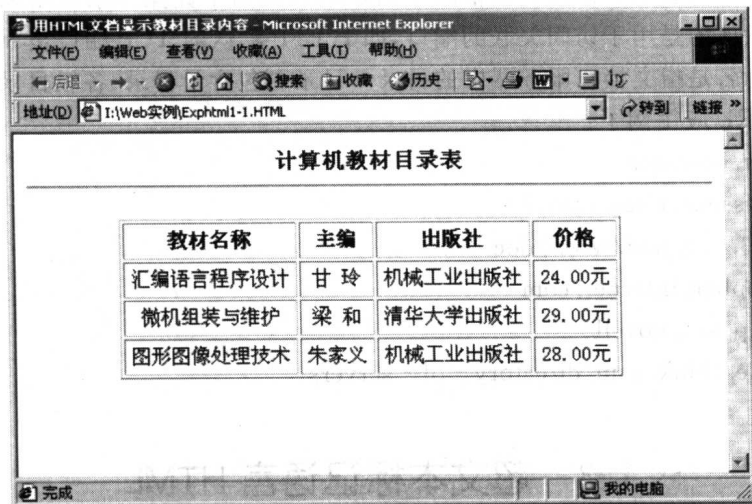


图 1-3 通过 Web 浏览器解释后呈现给用户的 Web 页面

通过上面的例子可以看出来,HTML 文档是一个呆板无生气的文字文档,Web 页面却又是

另一番让人赏心悦目的天地。由于 HTML 文档的编写不受操作系统平台的限制,只要书写成一个纯文本格式文件,然后用 .htm 或 .html 后缀保存起来即可。HTML 的优势有如下几点:

- (1) 每一个 HTML 文档占用空间小,易于传送。
- (2) HTML 文档是多平台兼容,只需有 HTML 浏览器就可在任何平台上阅读网页文件。
- (3) HTML 是标记性语言,不像其他一些程序设计语言受编程框架格式限制,易学易懂。
- (4) 只要有一个文本编辑软件即可编写 HTML 文档。当然专门的 HTML 编辑器生成 HTML 文档会更便捷,不需去记忆一些格式命令,点击工具或菜单即可完成。

HTML 语言也有它的一些局限性。如 HTML 的排版功能不像 Word 那样直观方便,达不到所见即所得的效果,要不断的通过浏览器来观察效果,视情况回过头来修改文档。另外,虽然 HTML 文档支持不同平台的浏览器,当不同的浏览器通过网络获取同一个 HTML 文档时,显示的结果可能不一样。即使用同一个浏览器,由于用户对浏览器的设置不同,同一个文档的阅读和显示形式也可能有差异。因此要求编制者在设计网页文件时,既要考虑不同形式的浏览器,又要尽量使文档结构清晰易懂。

有几点应引起 HTML 初学者的注意:

(1) HTML 文档的编辑工具:HTML 文档是纯文本文件,不含任何二进制控制码。因此可用任何一个纯文本编辑工具。如 Windows 系统下的记事本(NotePad)、Word, DOS 环境下的 Edit, UNIX 系统下的 Vi, emacs 等。

(2) 文件名的命名规则:由于 Internet 上的大多数 Web 服务器运行于 UNIX 环境下,对英文字母的大小写较敏感,所以一定区分大小写,同样一个文件以 htmdoc. html 和 HTMLDOC. HTML 命名,就是表示两个不同的文档。文件名中不能用空格和特殊字符。也尽量不要用汉字作文件名称部分。

(3) 由于编写出的 HTML 文档不是所见即所得,成功与否要通过浏览器查看效果,因此一个好的 HTML 文档,可能要经过多次的反复修改和显示才能成功。

1.3.1 HTML 的基本语法

网页是用 HTML 语言按一定语法规则编制的文本文档,通过浏览器解释并显示在用户浏览器的窗口中。用这些简单、易学的通用标记语言能够制作出包含文字、图像、声音、动画等内容丰富多彩的网页。

标记是 HTML 语言的标签符号和用标签标符号构成的各种元素的统称,用来界定各种单元,如标题、段落、列表等。标签是描述性的标记,用一对<>中间包含若干字符命令表示,通常成对出现。前一个“<>”为起始标记,后一个“</>”为结束标记,单元的名称和属性由起始标记给出。例如一段正文的表示如下:

<BODY> ;HTML 文档体起始标记

正文内容

</BODY> ;HTML 文档体结束标记

一般标记的书写不区分大小写,像上面的<BODY>与<body>视为相同。书中用大写字母表示标记符,为的是和文件内容区别开。但是 XHTML(即 HTML 和 XML 的混合主体)要求小写标记,小写标记可以使 HTML 文档向前兼容。“;”号及之后的文字描述是为了增强源文件

的可读性所加的注解,也可用“<!-- 注解内容 -->”作注解,浏览器并不解释它们。

标记是<>两符号之间的命令且成对出现(有个别例外)。学习 HTML 的过程也就是学习各种标记语法含义、属性的过程。

下面介绍 HTML 的基本标记。

1. <HTML>标记

<HTML>标记是 HTML 文件最外层的标记,即基本 HTML 页面的开始标记。在<HTML>和</HTML>之间是 HTML 文档的主体,包含 HEAD 标记和 BODY 标记等。<HTML>说明这个文件内容是用 HTML 语言实现的。HTML 文档的最后以</HTML>标记,说明 HTML 文档的结束。其语法规则是:

一般形式:<HTML>...</HTML>

属性:VERSION=String

说明

(1) 一般形式给出了一个标记通常的使用格式。

(2) 属性描述该标记的性质,VERSION 是可选的,属性修改标记的行为。很多标记都可以独立运行,但是有许多标记必须具有属性才能进行。属性只能应用到开始标记上,而不能在结束标记上放置属性。

(3) 值是属性的定义部分,最终要用值来修改标记。属性的值要用引号括起来。如<HTML>的属性值为 VERSION="//W3C//DTD HTML4.0 Final//EN"。DTD 定义了文档中有效的元素、属性及元素间的关系。该属性值表明,该 HTML 文档使用的 HTML 版本是万维网联盟(W3C)推行的 HTML 4.0 版本。

2. <HEAD>标记

该标记用来标记 HTML 文档的页首,提供有关文档本身的信息如网页标题、文档属性参数、使用的字符集、预设的 Script 语言、样式表定义等等不在页面上显示的网页元素。该标记的语法规则为:

一般形式:<HEAD>

:

</HEAD>

属性:None(无)

说明

介于<HEAD>和</HEAD>标记间的文本信息是为整个 HTML 文档的辅助说明,供阅读使用。一般<HEAD>标记间的标记文本力求简洁短小。

例 1-1 <TITLE>和</TITLE>标记的用法。

```
<HTML>
<HEAD>
<TITLE>本网页主题</TITLE>
</HEAD>
</HTML>
```