

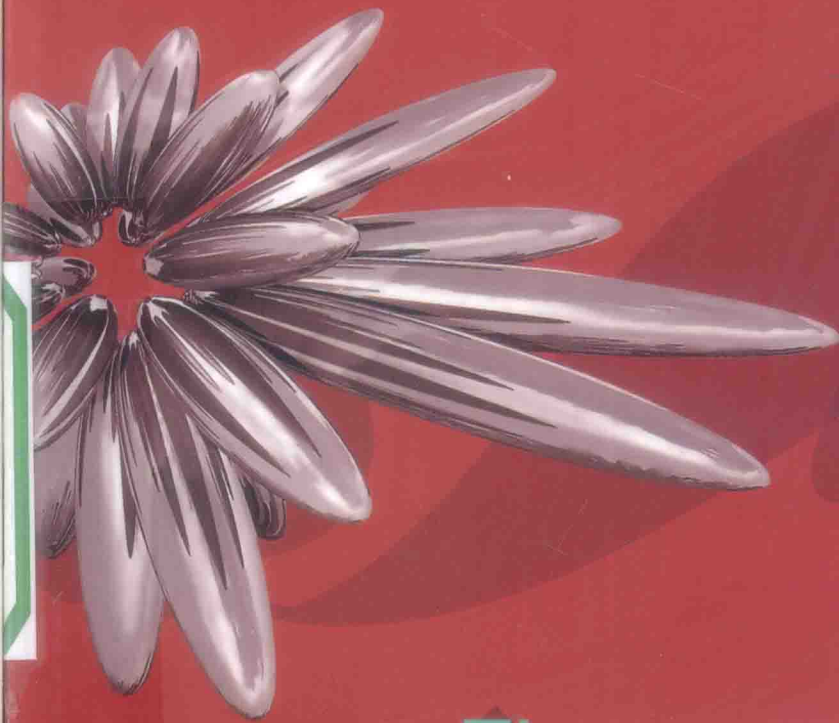


普通高等教育“十三五”规划教材

高等学校计算机规划教材

Web 技术基础

◆ 杨占胜 主编 ◆ 傅德谦 许作萍 副主编



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

普通高等教育“十三五”规划教材

Web 技术基础

杨占胜 主编

傅德谦 许作萍 副主编

孙晓燕 刘乃丽 参编



电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书首先从互联网与万维网的概念入手,介绍了 Web 在整个网络体系架构中的位置、Web 与 Internet 的关系,引出 Web 的三个基本要素: URL、HTTP、HTML。之后,全书系统、深入地介绍了 HTML 4.01、CSS 2.1、JavaScript 的全部内容,XML、XHTML、HTML 5、CSS 3 的基本原理与核心内容,以及 Web 编辑工具 Dreamweaver 的使用方法。按照知识关联和学习路线,全书对 Web 领域中的技术概念几乎都进行了介绍,并具有一定的深度和特色。对 Dreamweaver 软件的介绍,除其基本的功能外,着重于 Dreamweaver 最具特色与实用性的工具。

Web 技术涉及的计算机语言较多,内容庞杂,合理组织安排这些知识内容十分重要。本书的章、节、小节、知识点编排都是精心设计的,力求条理清晰、结构合理。Web 技术基础知识简单易懂,HTML、CSS 只有少量难点,JavaScript 有一定的难度。本书在介绍各知识点时,对简单的内容尽量简明扼要,对一些有难度的内容进行了深入细致的剖析。书中的例程是在多年教学实践中积累精选的,例程的代表性强,所有的例程组织链接成一个网站,可以作为 Web 前端设计技术人员的参考代码。书末附有实验指导书,8 个实验项目也可组织成一个小型的网站。本书配套教学资源有 PPT、源代码等。

对于没有基础的初学者,通过本书可以系统全面地学习 Web 技术的基础知识,对于有基础的技术人员,本书可以解答其一些常见的疑难问题,有助于其建立起系统的 Web 技术结构体系。本书可作为本、专科院校和各类培训学校计算机相关专业的教材,也适合网页设计、网站开发、Web 应用程序编程技术人员参考使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

Web 技术基础 / 杨占胜主编. —北京: 电子工业出版社, 2016.8
ISBN 978-7-121-29743-4

I. ①W… II. ①杨… III. ①网页制作工具—程序设计—高等学校—教材 IV. ①TP393.092.2

中国版本图书馆 CIP 数据核字 (2016) 第 200129 号

策划编辑: 任欢欢

责任编辑: 任欢欢

印 刷: 三河市鑫金马印装有限公司

装 订: 三河市鑫金马印装有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1092 1/16 印张: 25 字数: 640 千字

版 次: 2016 年 8 月第 1 版

印 次: 2016 年 8 月第 1 次印刷

定 价: 56.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: 192910558 (QQ 群)。

前 言

Web 赋予了互联网靓丽的青春和 Internet 强大的生命力,极大地推进了 Internet 的发展和普及。Web 改变了全球信息化的传统模式,带来了一个信息交流的全新时代。目前,Web 已成为人们共享信息的主要手段,是最流行的网络应用。Web 技术已成为计算机类专业最重要的学习内容之一。Web 技术基础知识虽然简单,但其内容庞杂,涉及标记语言、样式语言、脚本语言,语法各异,需要注意的细节多;其中的技术术语、概念也多;又有标准与非标准、版本更新、浏览器支持的差异。关于 Web 技术的书不少,但是能够将 Web 技术知识合理组织,既条理清晰又全面系统地介绍 Web 技术的书并不多。在多年的教学实践中,我们积累了一些课堂讲稿和大量的 Web 设计例程,希望编写一本结构合理、知识系统、讲解深入的 Web 技术基础教材。

本书首先从互联网、万维网的概念入手,介绍了 Web 在网络体系架构中的位置、Web 与 Internet 的关系,引出 Web 的三个基本要素:URL、HTTP、HTML。之后,全书系统、深入地介绍了 HTML 4.0.1、CSS 2.1、JavaScript 的全部内容,XML、XHTML、HTML 5、CSS 3 的基本原理与核心内容,以及 Web 编辑工具 Dreamweaver 的使用方法。按照知识关联和学习路线,全书对 Web 领域中的技术概念几乎都进行了介绍,如网络(Network)、互联网(Internet)、万维网(World Wide Web)、W3C(World Wide Web Consortium)、统一资源定位符(URL, Uniform Resource Locator)、统一资源命名(URN, Uniform Resource Name)、统一资源标识符(URI, Uniform Resource Identifier)、超文本传输协议(HTTP)、Web 浏览器、Web 服务器、多用途网际邮件扩展协议(MIME, Multipurpose Internet Mail Extensions)、网页(Web Page)、网站(Web Site)、主页(Home Page)、超文本标记语言(HTML)、层叠样式表(CSS)、JavaScript 脚本语言、DHTML(Dynamic HTML)、文档对象模型(DOM, Document Object Model)、可扩展标记语言(XML)、文档类型定义(DTD, Document Type Define)、可扩展样式语言(XSL, eXtensible Stylesheet Language)、Schema、XML DOM、XHTML、HTML5、CSS3、JSON(JavaScript Object Notation)、Web 1.0、Web 2.0、Web 3.0、Dreamweaver 等。

本书的章、节、小节、知识点编排都是精心设计的,力求条理清晰、结构合理。全书 10 章一气呵成,没有进行篇幅的划分,总体上分为两大部分,前 6 章为第一部分:Internet 与 Web、HTML、CSS、JavaScript、XML 简介、Web 编程工具;后 4 章为第二部分:HTML5、CSS3、JavaScript 进阶、Web 技术发展概述。第一部分的知识更为基础,第二部分是 Web 新技术和 JavaScript 中难度大的部分。

本书在介绍各知识点时,对简单的内容尽量简明扼要,对一些有难度的内容进行了深入细致的剖析。用了较多的篇幅阐述 HTML 中多媒体标记的使用;CSS 中 display、position、float、vertical-align 属性的功能细节;JavaScript 中浏览器对象与 HTML DOM 对象的关系;函数的多重功能、参数与伪继承;DOM 2 事件处理机制等复杂内容。对某些内容的解释有一定的深度和自己独特的见解。在介绍 Internet 和 Web 时,将其比喻为两张网。“在计算机网络技术中,通常会涉及两张网,即 Network 和 Web。Net 的原意为渔网,Network 主要指

硬件网络；Web 的原意为蜘蛛网，主要指软件网络。”同时，将浏览器定义为：“简单地说，浏览器是一个超文本文件解析程序，这个程序实现了 HTTP 协议”。将 HTML、CSS、XML 三者之间的关系描述为：“HTML 兼有语义和样式功能，但 HTML 的语义功能很弱，样式也不丰富；CSS 是对 HTML 样式的增强，而 XML 是在语义方面的补充和完善。”在介绍 CSS 中一些复杂样式时，一句话指出其主要作用与本质特征：“display 样式属性更改网页元素默认的盒模型类型”，“position 样式属性更改网页元素在正常流布局中的位置”。将正则表达式描述为：“正则表达式是通配符技术的扩展”，说明 XML 潜在的强大功能，“XML 文档同时具有面向对象技术、数据库技术、Web 技术三大功能特性：包含了计算机软件领域的主要技术点，体现了简单就是美（Simple is the best）的哲学思想。”将 JavaScript 函数描述为：“是普通函数、对象、方法、类”。对 Dreamweaver 的介绍，除其基本的功能外，着重于 Dreamweaver 最具特色与实用性的工具，如“标签选择器”、“批量修改”、“拖放链接”、“图像热点截取”、“插入 Flash 视频”等。书中的例程是在多年教学实践中积累精选的，例程的代表性强，所有的例程组织链接成一个网站，可以作为 Web 前端设计技术人员的参考代码。附录中的 8 个实验项目也可组织成一个小型的网站。

XML 作为 Web 技术的重要基础知识，不仅在 Web 方面应用广泛，在整个软件开发领域几乎无所不在，是计算机软件相关专业必备的基础知识。XML 涉及的内容很多，本书只对其基本原理和核心内容进行了介绍。有的学校可能专门开设 XML 课程，教师在使用本教材时可根据具体的教学计划决定是否讲授这一章。第 6 章 Dreamweaver 相关内容可根据情况分别在前面 4 章讲完后讲授一部分，第 2 章 HTML 相关内容讲完后，再讲授 Dreamweaver 的网页编辑功能；第 3 章 CSS 相关内容讲完后，再讲授 Dreamweaver 对样式表的支持；第 4 章 JavaScript 相关内容讲完后，再讲授 Dreamweaver 对 JavaScript 的支持；第 5 章 XML 相关内容讲完后，再讲授 Dreamweaver 对 XML 的支持。提前让学生使用 Dreamweaver 可加快课程的教学进度，但将 Dreamweaver 单独作为一章编写，使全书的结构更紧凑，系统性更强。第 2 章中的 HTML 标记都比较简单，建议在教学时按类型介绍一些常用的标记，其他的无须详细讲解，可由学生阅读教材自学。大部分学生都学习了 C 语言，或者 Java 语言，第 4 章中的运算符与表达式、流程控制与语句等 JavaScript 语法，在讲授时只需简略地提及一下。教材中这部分内容没有一笔带过，但尽量简明扼要，大部分语法知识采用表格列举，以便查阅参考；例程也较少，只有一些较典型的例子。书中未对例程的运行效果附图，例程只需用浏览器打开即可运行，建议在讲授时用多媒体设备直接演示，或者在机房由学生亲自运行，以增加学生的学习兴趣，提高教学效果。书末附有实验指导书，共 8 个实验项目，有 20 多个较典型的练习网页，可组织成一个小型的网站。为了方便教学，作者将提供本书所有例程源代码、实验项目源代码、PPT 等电子资源的下载。

本书由杨占胜主编和统稿。其中杨占胜编写了第 1 章至第 6 章、第 9 章和第 10 章、附录 A；杨占胜与许作萍共同编写了第 7 章；孙晓燕编写了第 8 章；刘乃丽副教授负责教材编写的组织工作，参与了第 4 章与部分习题的编写；傅德谦教授负责教材编写的指导工作，对本书的编写与出版工作进行了指导。电子工业出版社任欢欢编辑为本书的尽快出版付出了很多努力。由于本书涉及的知识点较多，加之编者水平有限，时间较紧，书中疏漏和不妥之处在所难免，真诚欢迎广大读者和同行批评指正，以使教材不断更新完善。作者 E-mail: 670418190@qq.com。

杨占胜

2016 年 5 月于山东临沂

Contents 目录

第1章 Internet 与 Web.....1	2.7.3 <tr>和<td>的属性.....34
1.1 互联网简介.....1	2.7.4 <table>的子元素.....37
1.2 万维网.....2	2.7.5 表格布局.....41
1.3 统一资源定位符.....3	2.8 表单标记.....45
1.4 超文本传输协议.....4	2.8.1 表单.....45
1.5 Web 浏览器.....6	2.8.2 输入控件.....46
1.6 Web 服务器.....7	2.8.3 列表控件.....47
1.7 资源类型标识.....8	2.8.4 文本域控件.....47
本章小结.....9	2.8.5 辅助标记.....49
思考题.....9	2.9 框架标记.....49
第2章 HTML.....10	2.9.1 框架集.....50
2.1 基本文档结构标记.....11	2.9.2 内联框架.....51
2.2 文本格式化标记.....12	2.10 元标记.....51
2.2.1 标题标记.....12	2.10.1 http-equiv 属性.....51
2.2.2 区段标记.....12	2.10.2 name 属性.....52
2.2.3 文字修饰标记.....13	2.10.3 content 属性.....52
2.2.4 特殊符号.....15	2.11 HTML 属性.....53
2.3 超链接标记.....16	2.11.1 必需属性.....53
2.3.1 链接地址.....16	2.11.2 通用属性.....53
2.3.2 链接标记的主要属性.....17	2.11.3 事件属性.....54
2.4 图像标记.....18	2.11.4 常用属性.....55
2.4.1 图像文件类型.....19	本章小结.....55
2.4.2 图像链接与图像映射.....20	思考题.....56
2.5 多媒体播放.....20	第3章 CSS.....57
2.5.1 嵌入标记和对象标记.....21	3.1 基本样式属性.....57
2.5.2 音频和视频文件格式.....26	3.1.1 字符样式属性.....57
2.6 列表标记.....30	3.1.2 文本样式属性.....58
2.6.1 无序列表.....30	3.1.3 背景样式属性.....60
2.6.2 有序列表.....31	3.1.4 边框样式属性.....61
2.6.3 定义列表.....31	3.1.5 边距样式属性.....63
2.7 表格标记.....32	3.1.6 列表样式属性.....65
2.7.1 表格.....32	3.2 选择器.....66
2.7.2 <table>的属性.....33	3.2.1 通配符选择器.....67

3.2.2 标记选择器	67	4.3.3 常量	121
3.2.3 类选择器	67	4.3.4 变量	124
3.2.4 id 选择器	67	4.3.5 常用全局函数	129
3.2.5 属性选择器	68	4.3.6 运算符与表达式	131
3.2.6 后代选择器	69	4.3.7 流程控制与语句	135
3.2.7 并列选择器	69	4.3.8 函数	141
3.2.8 子元素选择器	69	4.4 事件处理	143
3.2.9 相邻兄弟选择器	69	4.5 内置类	150
3.2.10 伪类选择器	69	4.5.1 数组 Array	150
3.2.11 伪元素选择器	72	4.5.2 日期 Date	152
3.3 网页中使用 CSS	74	4.5.3 数学 Math	153
3.3.1 内联样式表	74	4.5.4 字符串 String	155
3.3.2 内部样式表	74	4.6 HTML 文档对象	157
3.3.3 外部样式表	75	4.6.1 浏览器对象	158
3.4 定位相关属性	78	4.6.2 DOM 对象	168
3.4.1 盒模型与流布局	78	4.7 正则表达式	172
3.4.2 显示与大小属性	80	4.7.1 正则表达式的规则	172
3.4.3 定位与布局属性	83	4.7.2 常用正则表达式	174
3.4.4 内容修剪与对齐属性	91	4.7.3 JavaScript 使用正则表达式	175
3.5 其他样式属性	100	4.8 JavaScript 应用	176
3.5.1 表格相关属性	100	4.8.1 修改网页内容	176
3.5.2 鼠标样式属性	104	4.8.2 表单验证	180
3.5.3 轮廓相关属性	105	4.9 JavaScript 修改 CSS 样式	185
3.5.4 内容生成相关属性	106	本章小结	187
3.6 div + CSS 布局	110	思考题	188
本章小结	113	第 5 章 XML 简介	190
思考题	114	5.1 XML 概述	190
第 4 章 JavaScript	115	5.2 XML 语法	192
4.1 JavaScript 概述	115	5.2.1 XML 语法规则	192
4.1.1 JavaScript 的特点	115	5.2.2 XML 语法元素	193
4.1.2 Java 与 JavaScript 的区别	116	5.2.3 格式良好和有效的 XML 文档	194
4.1.3 两个简单的输出方法	117	5.3 DTD	194
4.2 网页中嵌入 JavaScript	117	5.3.1 元素类型声明	195
4.3 JavaScript 语法	119	5.3.2 属性类型声明	196
4.3.1 基础语法点	120	5.3.3 实体	198
4.3.2 基本数据类型	120	5.3.4 XML 中声明 DTD	201

5.4 名称空间	203	6.3.5 多媒体	242
5.5 Schema	204	6.3.6 表格	243
5.5.1 Schema 概述	204	6.3.7 表单	245
5.5.2 Schema 的定义	205	6.3.8 框架	245
5.5.3 数据类型	208	6.4 DOCTYPE 声明与网页解析	
5.5.4 简单类型定义	210	模式	246
5.5.5 复杂类型定义	212	6.4.1 网页文档类型声明	246
5.5.6 XML 中声明 Schema	214	6.4.2 浏览器的工作模式	248
5.6 CSS 格式化 XML	214	6.5 网页布局	249
5.7 XSL	216	6.6 网站模板	250
5.7.1 XSL 概述	216	6.7 CSS 的支持	251
5.7.2 XSLT 文档结构	217	6.8 JavaScript 的支持	253
5.7.3 XSLT 模板	217	6.9 XML 的支持	254
5.7.4 模式处理	219	6.10 参考资源	254
5.7.5 节点选择	220	本章小结	255
5.8 XML 解析器	222	思考题	256
5.9 XML DOM	223	第 7 章 HTML 5	257
5.9.1 Document 文档对象	223	7.1 HTML 5 概述	257
5.9.2 Node 节点对象	225	7.1.1 从 HTML 到 XHTML	
5.9.3 NodeList 节点列表对象	225	和 HTML 5	257
5.9.4 NamedNodeMap 无序节点		7.1.2 HTML 5 的优势	258
集对象	226	7.2 HTML 5 新增常用元素和	
5.9.5 DOM 例程	226	属性	259
5.10 XHTML	229	7.2.1 新增的文档结构元素	259
本章小结	230	7.2.2 新增的通用属性	262
思考题	231	7.2.3 其他元素	265
第 6 章 Web 编程工具	232	7.3 HTML 5 增强的表单功能	266
6.1 Dreamweaver 界面	232	7.3.1 新增的表单元素和属性	266
6.2 站点管理	234	7.3.2 <input> 元素 type 属性	
6.2.1 站点建立	234	新增的属性值	270
6.2.2 文件管理	235	7.3.3 新增的客户端校验属性	273
6.2.3 资源管理	235	7.3.4 增强的文件上传域	273
6.2.4 站点地图	236	7.4 多媒体播放	276
6.3 网页编辑	236	7.4.1 音频和视频标记	276
6.3.1 编码辅助功能	237	7.4.2 JavaScript 脚本控制	
6.3.2 可视化编辑	238	媒体的播放	278
6.3.3 超链接	239	7.5 拖放行为	280
6.3.4 图像	240	7.5.1 拖放 API	280

7.5.2 拖放操作	281
7.6 绘图功能	283
7.6.1 绘图 API	283
7.6.2 绘制图形	285
本章小结	293
思考题	294
第 8 章 CSS 3	295
8.1 CSS 3 新增的选择器	295
8.1.1 伪类选择器	295
8.1.2 兄弟选择器	300
8.1.3 浏览器前缀	300
8.2 服务器字体	301
8.2.1 @font-face	301
8.2.2 服务器字体与客户端 字体结合使用	303
8.3 边框和阴影	303
8.3.1 渐变边框	303
8.3.2 圆角边框	304
8.3.3 图片边框	305
8.3.4 阴影	306
8.4 分栏布局	306
8.4.1 分栏属性	307
8.4.2 分栏显示	308
8.4.3 块框属性	310
8.5 变形与动画	310
8.5.1 变形	310
8.5.2 Transition 动画	312
8.5.3 Animation 动画	314
本章小结	317
思考题	317
第 9 章 JavaScript 进阶	319
9.1 JavaScript 函数高级功能	319
9.1.1 函数定义	319
9.1.2 函数的特性	322
9.1.3 类属性	323
9.1.4 函数的调用	324
9.1.5 函数的独立性	326

9.1.6 函数的参数	327
9.1.7 类的扩展	331
9.1.8 对象的创建	333
9.2 DOM 事件模型	335
9.2.1 基本事件模型	336
9.2.2 DOM 2 事件模型	342
9.2.3 IE 事件模型	351
9.3 JavaScript 程序调试	356
9.3.1 显示脚本错误	356
9.3.2 Firebug	357
本章小结	362
思考题	362
第 10 章 Web 技术发展概述	364
10.1 Web 技术发展的阶段	364
10.1.1 Web 技术发展的三个 阶段	364
10.1.2 Web 应用程序发展的三个 阶段	365
10.2 Web 2.0 概述	366
10.2.1 Web 1.0 的特征	366
10.2.2 Web 2.0 的特征	367
10.2.3 Web 2.0 应用技术	368
本章小结	369
思考题	369
附录 A	370
实验说明	370
实验 1 基本网页设计	373
实验 2 CSS 应用	377
实验 3 JavaScript 编程	379
实验 4 表单验证	382
实验 5 XML 文档操作	384
实验 6 HTML 5 应用	386
实验 7 DOM 2 事件处理	388
实验 8 个人网站设计	390
参考文献	391

计算机与网络的出现是继语言、文字、造纸与印刷术、电报与电话之后的第五次信息技术革命,计算机与网络已经深入到社会生活的各个方面。在计算机网络技术中,通常会涉及两张网——Network 和 Web。Net 的原意为渔网,Network 主要指硬件网络,它包括了 TCP/IP (Transmission Control Protocol/Internet Protocol) 四层网络体系结构中的三层,即主机-网络层、互联层、传输层,是由实现了这些层级协议的硬件设备,如网线、网卡、中继器、集线器、交换机、路由器、计算机(安装网络操作系统,即实现 TCP/IP 的操作系统)等组成的。Web 的原意为蜘蛛网,它是 World Wide Web (WWW) 的简称,中文称为万维网,主要指软件网络。Web 是 TCP/IP 网络体系结构中最顶层——应用层的应用之一。应用层中有远程登录(Telnet)、文件传输(FTP)、Web 应用(HTTP)、电子邮件(SMTP 和 POP3)等应用,在计算机网络发展的早期,Telnet 和电子邮件曾先后是最广泛的网络应用,Web 技术出现后即呈现出爆炸式的发展趋势,目前 Web 已成为人们共享信息的主要手段,是最流行的网络应用,其他的网络应用也都趋向于通过 Web 来实现。就像网络的发展和普及形成计算机即网络的概念一样,WWW 几乎成为 Internet 的代名词。但互联网(Internet)并不等同于万维网(Web),Web 只是互联网上基于超文本的软件应用系统。

1.1 互联网简介

计算机网络是使用通信线路和遵循特定技术规范(即实现特定网络协议)的设备将多台计算机连接起来,并能利用软件实现通信和资源共享的系统。计算机网络按照连接设备的范围分为局域网、城域网和广域网。局域网与广域网在地理范围上的巨大差异导致其通信方式的不同,局域网的通信方式为共享网络,广域网的通信方式为点对点。网络体系结构及其通信协议和共享网络通信是局域网技术的核心内容,网络互连是广域网技术的核心内容,这些都是计算机网络课程的范畴,而点对点通信技术是通信专业的主要内容。以太网是应用最广泛的局域网。局域网使用通信线路和网络设备(一般为路由器)连接起来就成为了广域网。城域网介于局域网与广域网之间,以提供接入、媒体、信息服务为主,强调业务功能和服务质量。目前城域网仍没有完全统一的技术标准。全世界范围内的局域网、城域网、广域网互相连接而成的一个超级广域网就是 Internet,称为互联网、因特网,Internet 是网络的网络。

因特网使用 TCP/IP 协议让不同的设备可以彼此通信。但使用 TCP/IP 协议的网络并不一定是因特网,一个局域网也可以使用 TCP/IP 协议。互联网的应用方式早期主要基于大

型主机系统，目前采用客户机/服务器体系结构。小型网络中的对等网络，各台计算机既可以作为服务器，也可以作为客户机，可身兼二职，是一种特殊的客户机/服务器结构模式。Web 应用基于 Internet，其安全机制，运行、访问方式都以 Internet 环境为基础，Web 技术中的许多主题与 Internet 相关。Web 应用的应用层协议为 HTTP (Hyper Text Transfer Protocol)，底层通常采用 TCP/IP 协议。Web 应用是典型的客户机/服务器结构，服务器程序称为 Web 服务器，或 HTTP 服务器，客户端程序采用统一的浏览器，所以又称为 B/S (Browser/Server) 应用程序。

1.2 万维网

万维网，即 World Wide Web，简称 WWW 或 Web。万维网是互联网的功能之一。万维网是建立在客户机/服务器模型之上，以超文本标记语言 (HTML, Hypertext Markup Language) 与超文本传输协议 (HTTP) 为基础，提供面向 Internet 服务的、用户界面一致的主从结构分布式超媒体系统。其中的服务器为 Web 服务器，又称 WWW 服务器。Web 服务器提供由 HTML 编写的超文本文件 (称为网页) 和其他资源，这些网页及资源采用超链接互相连接，它们既可放置在同一台服务器上，也可放置在不同地理位置的服务器上。用户使用统一的客户端软件 (浏览器) 浏览服务器上的网页和资源，客户端浏览器的主要功能是向服务器发送 HTTP 请求并显示服务器返回的 HTTP 应答信息页。万维网的工作方式如图 1-1 所示。



图 1-1 Web 工作方式

万维网起源于欧洲粒子物理研究室 CERN (European Organization for Nuclear Research)。1989 年 3 月，欧洲粒子物理研究室的 Tim Berners-Lee (蒂姆·贝纳斯·李) 提出 Web 计划。1980 年代后期，超文本技术已经出现，但没有人能想到把超文本技术应用到计算机网络上来，超文本只是一种新型的文本而已。Tim 创造性地将超文本技术应用于计算机文件中，通过在文件中嵌入超链接，把网络上不同计算机 (主机) 上的信息连接起来，并且通过超文本传输协议 (HTTP) 在 Web 服务器和客户机之间传输。Tim 成功开发出了世界上第一个 Web 服务器 (httpd) 和第一个 Web 客户端浏览编辑程序 (World Wide Web)。1989 年 12 月，他的发明正式定名为 World Wide Web，即 WWW。1991 年 8 月 6 日，Tim 建立的第一个网站 (也是世界上第一个网站) <http://info.cern.ch/> 上线。Web 在互联网上的出现，立即引起轰动，获得了极大的成功，并被迅速推广应用。Web 出现之前，互联网上内容的表现形式极其单调枯燥，资源的访问权限也很严格，网络操作较为复杂。而 Web 服务器能发布图文并茂的信息，甚至还可以发布音频和视频信息。人们只要通过使用简单的方法，就可以很迅速方便地获取丰富的信息资料。此外，Internet 的许多其他功能，如 E-mail、Telnet、FTP 等都可以通过 Web 实现。Web 是 Internet 上取得的最激动人心的成就，为 Internet 的普及迈出了开创性的一步。Web 技术给 Internet 赋予了强大的生命力，Web 浏览的方式带给了互联网靓丽的青春。Web 技术的发明改变了全球信息化的传统模式，开启了一个信息交流的全新时代。

Tim 在发明万维网之后,又相继制定了互联网的 URI、HTTP、HTML 等技术规范。1994 年,他在美国麻省理工学院成立了非营利性的万维网联盟 W3C(World Wide Web Consortium),邀请 Microsoft、Netscape、Sun、Apple、IBM 等共 155 家互联网著名公司,致力于 WWW 技术协议的标准化,进一步推动 Web 技术的发展。W3C 对互联网技术的发展和應用起到了基础性和根本性的支撑作用,是 Web 技术领域最具权威和影响力的国际中立性技术标准机构。Tim 被业界公认为“互联网之父”。

Web 是互联网上由许多互相链接的超文本组成的系统,在这个系统中,每个信息单元称为一个“资源”,由一个全局“统一资源定位符”(URL)来定位;这些资源主要是 HTML 编写的超文本文件(网页);通过 HTTP 传送给用户。Web 的三个基本要素分别是:资源寻址定位的统一资源定位器(URL);资源访问方式的超文本传输协议(HTTP);以及资源表达编辑的超文本标记语言(HTML)。

➔ 1.3 统一资源定位符

统一资源定位符(URL, Uniform Resource Locator),又称为统一资源定位器,是 Web 资源的全局地址表示,用于在互联网上寻址定位 Web 资源。与 URL 类似的还有一个统一资源标识符(URI, Uniform Resource Identifier),用来在互联网上唯一地标识一个资源。URL 是一种具体的 URI,URL 不仅可以用来标识一个资源,同时还能够定位这个资源。另外与 URL 并列的一个术语是统一资源命名(URN, Uniform Resource Name),它通过名字来标识资源,比如 mailto:java-net@java.sun.com。概括地说,URI 是以一种抽象的高层次概念定义统一资源标识,它可以是相对的;而 URL 和 URN 则是具体的资源标识方式,它们一般是绝对的。URL 和 URN 都是一种 URI。

URL 的格式为: protocol://hostname(or IP):[port]/website/path/[file][?query][#fragment]

协议://主机域名(或 IP 地址):端口号/网站/目录/文件名?查询参数#信息片断

例如: http://www.lyu.edu.cn:80/chpage/index.html?str=abc#a1

URL 主要由四部分组成:

第一部分是协议(protocol),或称为服务方式,主要的协议或者服务方式为 http、ftp、file、gopher、https、mailto、news。第一部分和第二部分之间用“://”符号隔开。

第二部分是服务器的域名(或 IP 地址)与端口号,域名是 IP 地址的文字表示,通过 DNS 进行转换。主机域名组成形式通常为:机器名+域名+域树+域林。HTTP 的默认端口号为 80,URL 中可以省略此默认的端口号。常用协议的默认端口号分别是 telnet: 23; ftp: 21; smtp: 25; pop3: 110; dns: 53。

第三部分是主机上资源的具体地址,如网站名(或虚拟目录名)、目录和文件名等。第二部分和第三部分用“/”符号隔开。

第四部分是参数,HTTP 请求时可以向服务器传递参数,参数格式为: ?名 1=值 1&名 2=值 2,与第三部分以“?”分隔,参数之间以“&”符号分隔。

另外,在同一网站内寻址,URL 的后面还可以加上网页锚点,网页锚点是页面内部信息片断的地址,是页面内部的定位寻址机制。

➔ 1.4 超文本传输协议

HTTP 是 Hypertext Transfer Protocol 的缩写, 即超文本传输协议。顾名思义, HTTP 提供了访问超文本信息的功能, 是 Web 浏览器和 Web 服务器之间的应用层通信协议。Web 使用 HTTP 协议传输各种超文本页面和数据。HTTP 可以使浏览器更加高效, 使网络传输减少。它不仅保证计算机正确快速地传输超文本文档, 还确定传输文档中的哪一部分, 以及哪部分内容首先显示 (如文本先于图形) 等。HTTP 包含命令和传输信息, 不仅可用于 Web 访问, 也可以用于其他因特网/内联网应用系统之间的通信, 从而实现各类应用资源超媒体访问的集成。

HTTP 协议底层通常采用 TCP/IP 协议, 但它并没有规定必须使用和基于的支持层。事实上, HTTP 可以在任何其他互联网协议上, 或者在其他网络上实现。HTTP 只假定其下层协议提供可靠的传输, 任何能够提供这种保证的协议都可以被其使用。

基于 HTTP 协议的客户端/服务器模式会话 (信息交换) 分四个过程: 建立连接、发送请求信息、发送响应信息、关闭连接。

(1) 建立连接: 客户端的浏览器向服务器端发出建立连接的请求, 服务端给出响应就可以建立连接了。

(2) 发送请求: 客户端按照协议的要求通过连接向服务器端发送自己的请求。

(3) 给出应答: 服务器端按照客户端的要求给出应答, 把结果 (HTML 文件) 返回给客户端。

(4) 关闭连接: 客户端接到应答后关闭连接。

HTTP 协议采用了请求/响应模型。客户端向服务器发送一个请求, 请求头包含请求的方法、URL、协议版本, 以及包含请求修饰符、客户信息和内容的类似于 MIME 的消息结构。服务器以一个状态行作为响应, 响应的内容包括消息协议的版本, 成功或者错误编码加上包含服务器信息、实体元信息以及可能的实体内容。通常 HTTP 消息有请求和响应两种类型。这两种类型的消息都由一个起始行、一个或者多个头域、一个指示头域结束的空行和可选的消息体组成。

HTTP 请求的格式:

```
HTTP 请求方法 请求的目标资源 HTTP/版本号
HTTP 头域
空行
HTTP 消息主体
```

简单的请求消息示例:

```
GET/chpage/index.html.HTTP/1.1
Host:www.lyu.edu.cn
Cache-Control: no-cache

<h1>Hello</h1>
```

HTTP 请求方法包括 GET、HEAD、POST、OPTIONS、PUT、DELETE、TRACE。方法 GET 和 HEAD 应该被所有的通用 WEB 服务器支持, 其他方法的实现是可选的。GET 方

法用于获取请求的目标资源。HEAD 方法也是获取请求的目标资源,只是可以在响应时,不返回消息体。POST 方法请求服务器接收包含在请求中的实体信息,可以用于提交表单,向新闻组、BBS、邮件群组 and 数据库发送消息。OPTIONS 方法返回服务器针对特定资源所支持的 HTTP 请求方法,可以利用该请求方法来测试服务器的功能性。PUT 方法向指定资源位置上传其最新内容。DELETE 方法请求服务器删除目标资源。TRACE 方法回显服务器收到的请求,主要用于测试或诊断。

浏览器中最常用的请求方式是 GET 和 POST。直接在浏览器地址栏输入访问地址所发送的请求,或提交表单发送请求时,没有设置 method 属性,或设置 method 属性为“get”,这几种都是 GET 方式的请求。GET 方式的请求会将请求参数的名和值转换成字符串,并附加在原 URL 之后,因此可以在地址栏中看到请求参数名和值。且 GET 请求传送的数据量较小,一般不能大于 2KB。POST 方式的请求通常需要设置 form 元素的 method 属性为“post”。传送的数据量较大,用户不能在地址栏里看到请求参数值,安全性相对较高。

HTTP 请求中的头域包括通用头域、请求头域和实体头域三个部分。每个头域由一个域名,冒号(:)和域值三部分组成。域名是大小写无关的,域值前可以添加任何数量的空格符,头域可以被扩展为多行,在每行开始处,使用至少一个空格或制表符。

通用头域包含请求和响应消息都支持的头域,通用头域包含 Cache-Control、Connection、Date、Pragma、Transfer-Encoding、Upgrade、Via。对通用头域的扩展要求通信双方都支持此扩展,如果存在不支持的通用头域,一般将会作为实体头域处理。

请求头域可能包含下列字段 Accept、Accept-Charset、Accept-Encoding、Accept-Language、Authorization、From、Host、If-Modified-Since、If-Match、If-None-Match、If-Range、If-Range、If-Unmodified-Since、Max-Forwards、Proxy-Authorization、Range、Referer、User-Agent。对请求头域的扩展要求通信双方都支持,如果存在不支持的请求头域,一般将会作为实体头域处理。

请求和响应消息都可以包含实体头域,实体头域包含关于实体的原信息,实体头域包括 Allow、Content-Base、Content-Encoding、Content-Language、Content-Length、Content-Location、Content-MD5、Content-Range、Content-Type、Etag、Expires、Last-Modified、Extension-Header。Extension-Header 允许客户端定义新的实体头,但是这些域可能无法被接收方识别。

HTTP 响应的格式:

```
HTTP/版本号 状态代码 状态描述
响应头部字段
空行
响应主体
```

简单的响应消息示例:

```
HTTP/1.1 200 OK
Content-Type: text/plain
```

```
<h1>Hello World!</h1>
```

状态代码主要用于机器自动识别,状态描述主要用于帮助用户理解。状态代码由三个数字组成,第一个数字定义响应的类别,后两个数字没有分类的作用。第一个数字可能取 5 个

不同的值:

- 1xx: 信息响应类, 表示接收到请求并且继续处理;
- 2xx: 处理成功响应类, 表示动作被成功接收、理解和接受;
- 3xx: 重定向响应类, 为了完成指定的动作, 必须接受进一步处理;
- 4xx: 客户端错误, 客户端请求包含语法错误或者是不能正确执行;
- 5xx: 服务器端错误, 服务器不能正确执行一个正确的请求。

404 Not Found 是一个最为常见的状态码, 它意味着无法找到请求的资源, 通常是 URL 不正确。200 OK 是用户想要得到的, 它意味着请求被正确处理, 没有什么错误。500 Server Error 则意味着服务器遇到了问题, 不能完成请求, 通常是动态页面中的代码发生错误。

HTTP 响应中的头域包括通用头域、响应头域和实体头域三个部分。通用头域和实体头域类似于 HTTP 响应。响应头域包含 Age、Location、Proxy-Authenticate、Public、Retry-After、Server、Vary、Warning、WWW-Authenticate。对响应头域的扩展要求通信双方都支持, 如果存在不支持的响应头域, 一般将会作为实体头域处理。

超文本传输协议已经演化出了很多版本, 它们中的大部分都是向下兼容的。客户端在请求的开始告诉服务器它采用的协议版本号, 而后者则在响应中采用相同或者更早的协议版本。HTTP/0.9 已过时, 因为它只接受 GET 一种请求方法, 没有在通信中指定版本号, 且不支持请求头。由于该版本不支持 POST 方法, 所以客户端无法向服务器传递太多信息。

HTTP/1.0 这是第一个在通信中指定版本号的 HTTP 协议版本, 至今仍被广泛采用, 特别是在代理服务器中。

HTTP/1.1 是当前版本, 持久连接被默认采用, 并能很好地配合代理服务器工作, 还支持以管道方式同时发送多个请求, 以便降低线路负载, 提高传输速度。

➡ 1.5 Web 浏览器

浏览器是 Web 系统中的客户机程序, 用户利用浏览器来访问服务器中的网页等资源。简单地说, 浏览器是一个超文本文件解析程序, 这个程序实现了 HTTP 协议。最初的浏览器是基于文本的, 不能显示任何类型的图形信息, 也没有图形用户界面 (GUI, Graphical User Interface), 这在很大程度上限制了 Web 的使用。1993 年, 美国伊利诺伊大学 (University of Illinois) 的国家超级计算机应用中心 (NCSA, National Center for Supercomputing Applications) 发布了具有图形用户界面的浏览器 Mosaic。普通用户借助于这个工具可以方便地访问 Web, Mosaic 带来的这种强大功能和便利, 直接导致了 Web 应用的爆炸性增长。

1994 年 4 月, Mosaic 的核心开发人员 Marc Andreessen (马克 安德森) 和 Jim Clark (吉姆 克拉克, 硅图 SGI, Silicon Graphics Incorporation 创始人) 在美国加州设立了 “Mosaic Communication Corporation”。由于伊利诺伊大学拥有 Mosaic 的商标权, 且已将其转让给望远镜娱乐公司 (Spyglass Entertainment), 因此 Andreessen 开发团队彻底重新撰写浏览器程序的源代码, 并将浏览器名称更改为 “Netscape Navigator”, 公司名字也改为 “Netscape Communication Corporation”, 中译为 “网景”。网景公司获得了巨大的成功, Navigator 浏览器成为当时最热门、市场占有率为首位的浏览器。

1995年,微软向 Spyglass 公司买下 Mosaic 的授权,以此为基础开发了 Internet Explorer,进军浏览器市场。之后网景公司的 Netscape Navigator 与微软公司的 Internet Explorer 之间进行了“浏览器大战”。1997年,微软推出功能更先进的 IE 4.0。1998年,微软发布 Windows 98 操作系统时捆绑 IE 5,这一系列技术攻势和市场策略使网景公司 Netscape Navigator 市场占有率从 90%急剧下降,最后被美国在线(AOL)收购。

1998年,网景公开了它的浏览器源码,并重新命名为 Mozilla,全部程序进行了重写,2002年发布了第一个版本。2004年,基于 Mozilla 源码的 Firefox 发布。Firefox 浏览器很成功,不断地蚕食微软的 IE 浏览器市场。目前市场上的主流浏览器有谷歌公司(Google)的 Chrome,挪威欧普拉软件公司(Opera Software ASA)的 Opera,苹果公司(Apple)的 Safari, Mozilla 基金会的 Firefox,微软的 IE (Internet Explorer)。

➡ 1.6 Web 服务器

Web 服务器本质上是驻留于因特网上某种类型计算机的程序,通常将运行 Web 服务器程序的计算机称为 Web 服务器。Web 服务器一般具有独立的 IP 地址,当 Web 浏览器(客户端)连到服务器上并发送 HTTP 请求时,服务器将处理该请求并将请求的资源以 HTTP 应答返回到该浏览器上。Web 服务器是另一个实现 HTTP 协议的程序,它的基本功能是响应和处理 HTTP 请求,即接收、解析、应答客户端的 HTTP 请求。Web 服务器的工作原理并不复杂,可分成 4 个步骤:连接过程、请求过程、应答过程以及关闭连接。服务器是一种被动程序,只有当 Internet 上其他计算机发出请求时,服务器才会响应。

传统 Web 服务器的作用是整理和储存各种 WWW 资源,响应客户端软件的请求,把用户所需的资源传送到客户端机器上。目前大部分 Web 服务器除能完成这些基本功能外,还能运行网页中嵌入的服务器端脚本程序,这样的 Web 服务器又称为应用程序服务器,应用程序服务器是 Web 服务器的扩展,Web 服务器可当成应用程序服务器的一个子集。HTML 文件中嵌入的服务器端脚本程序为 VB 代码的是 ASP 网页;服务器端脚本程序为 Java 代码的是 JSP 网页;服务器端脚本程序为 PHP 代码的是 PHP 网页。服务器端脚本程序通常具有事务处理(Transaction Processing),数据库连接(Database Connectivity)和消息传输(Messaging)等功能。服务器端脚本程序的强大功能使得 Web 应用可以随心所欲地实现各种复杂的商业逻辑。目前 Web 服务器成为 Internet 上最大的计算机群,Web 文档之多、功能之强大、链接的网络之广泛,令人难以想象。

在 UNIX 和 Linux 平台下使用最广泛的 Web 服务器是 Apache,而 Windows 平台使用的 Web 服务器为 IIS (Internet Information Server),其他著名的 Web 服务器还有 IBM 的 WebSphere、美国在线的 AOL Server、W3C 组织的 Jigsaw、俄罗斯的 Nginx。Apache 与 Nginx 是纯静态网页服务器,支持服务器端动态程序需要外部的功能模块,IIS 是一个强大的应用程序服务器,支持 ASP 和 ASP.NET 动态技术,其余几个都是基于 Java 的应用程序服务器。在选择使用 Web 服务器时应考虑的因素有:性能、安全性、日志和统计、虚拟主机、代理服务、缓冲服务和集成应用程序等。

Apache 是世界上用得最多的 Web 服务器,市场占有率达 60%左右。它源于 NCSA 的 httpd

服务器，当 NCSA WWW 服务器项目停止后，那些使用 NCSA WWW 服务器的人们开始交换用于此服务器的补丁，这也是 Apache 名称的由来（a patch 补丁的异体）。世界上很多著名的网站都是 Apache 的产物，它的成功之处主要在于它的源代码开放，有一支开放的开发队伍，支持跨平台的应用（可以运行在几乎所有的 UNIX、Linux、Windows 系统平台上）以及它的可移植性等方面。

Microsoft 的 Web 服务器产品为 Internet Information Services (IIS)，IIS 是允许在公共 Intranet 或 Internet 上发布信息的 Web 服务器。IIS 是目前最流行的 Web 服务器产品之一，很多著名的网站都是建立在 IIS 平台上的。IIS 提供了一个图形界面的管理工具，称为 Internet 服务管理器，可用于监视配置和控制 Internet 服务。

➡ 1.7 资源类型标识

电子邮件初期只能传送 US-ASCII 字符的文本，为了在邮件中包含非 US-ASCII 字符的文本，以及图像、音频、视频等媒体内容，互联网应用开发者采取了多种方法，多用途网际邮件扩充协议（MIME，Multipurpose Internet Mail Extensions）设计的目标就是允许各种不同类型的文档都可能通过 Internet 邮件进行发送。MIME 通过统一的标准定义邮件内容的信息格式，使邮件发送者、邮件代理、邮件接收者能够正确地解析邮件的内容，从而达到传送多媒体的通信目的。目前 MIME 是解决邮件中包含复合型数据问题的成熟机制，而且被用于互联网上标注资源的类型。Web 中也采用了 MIME 来指定所传递的文档类型。Web 服务器在一个将要发送到浏览器的文档头部附加了 MIME 的格式说明。当浏览器从 Web 服务器中接收到这个文档时，它就根据其中包含的 MIME 格式说明来确定下一步的操作，比如，如果文档内容为文本，则 MIME 格式说明将通知浏览器文档的内容是文本，并指明具体的文本类型；如果文档内容为音频，那么 MIME 格式说明将通知浏览器文档的内容是音频，并给出音频格式的具体描述，以便于浏览器选用合适的播放程序，正确播放接收的音频文件。

MIME 信息类型（Content-Type）定义的基本格式为：媒体类型/子类型，媒体类型被用来声明数据的一般类型，而子类型则指明了数据的细节格式。顶层媒体类型有 7 种：文本媒体类型（text）、图像媒体类型（image）、音频媒体类型（audio）、视频媒体类型（video）、应用媒体类型（application）、多部分媒体类型（multipart）、报文媒体类型（message）。其中前 5 个为离散媒体类型，后 2 个为复合媒体类型。媒体类型值可能会扩张，通过扩张，子类型集合不断增长。为了确保媒体类型的子属性值发展有序、规范、风格一致，MIME 媒体类型机制建立了一个登记程序，使用 IANA（Internet Assigned Numbers Authority）作为 MIME 各个方面扩展的中心登记处，见表 1-1。

表 1-1 常见 MIME 类型

类型/子类型	扩展名	类型/子类型	扩展名
text/plain	.txt(.c .h)	application/octet-stream	.bin(.class .exe)
text/html	.html(.htm .stm)	application/postscript	.ps(.ai .eps)
text/css	.css	application/hta	.hta