

Web

应用开发原理与技术



王继成 武港山 等编著

 **机械工业出版社**
CHINA MACHINE PRESS



Web 应用开发原理与技术

王继成 武港山 等编著



机械工业出版社

作为计算机网络技术应用的热点之一，Web 应用开发吸引了大量用户的兴趣。本书以“基本原理”和“关键技术”为重点，分五个部分对 Web 应用开发进行讨论。第一部分简要介绍了 Web 应用开发中涉及的各方面技术内容；第二部分讲述 Web 上客户与服务器之间的交互协议 HTTP，以及基于 HTTP 的应用开发；第三部分讲述 Web 上数据表示规范 HTML，特别是与应用开发相关的表单、脚本、样式表等内容；第四部分简要介绍了脚本语言，并讨论了客户端应用开发框架 DHTML 以及脚本在客户端的应用；第五部分讲述 Web 服务器端 CGI 应用程序和 ASP 脚本程序的开发，以及服务器端状态信息的维护。本书最后还附加了一个专题讨论，对 Web 信息检索和挖掘这一热门应用进行了介绍。

本书结构合理、内容紧凑、示例典型，既可作为高等院校计算机及相关专业的教材，也可供对 Web 应用开发感兴趣的读者自学，或者作为相关技术人员的参考书。

图书在版编目（CIP）数据

Web 应用开发原理与技术/王继成等编著. —北京：机械工业出版社，2003.5

ISBN 7-111-11929-0

I. W... II. 王... III. 计算机网络—基本知识 IV. TP393

中国版本图书馆 CIP 数据核字（2003）第 024381 号

机械工业出版社（北京市百万庄大街 22 号 邮政编码 100037）

策 划：胡毓坚

责任编辑：田 梅

责任印制：闫 焱

北京京丰印刷厂印刷·新华书店北京发行所发行

2003 年 4 月第 1 版·第 1 次印刷

787mm×1092mm 1/16·18.5 印张·454 千字

0 001—5 000 册

定价：28.00 元

凡购本图书，如有缺页、倒页、脱页，由本社发行部调换

本社购书热线电话（010）68993821、88379646

封面无防伪标均为盗版

前 言

■ 背景

目前, Internet 已经成为全球信息高速公路的基础, 而 Web 是 Internet 上使用最多的服务之一。在 Web 诞生的初期, Web 只是扮演着信息发布平台的角色, 人们可以在 Web 上快速有效地发布和获取信息, 不受时空的限制。而现在, Web 已逐渐地发展成为 Internet 应用开发的平台, 在此基础上, 电子商务、远程教育等新型服务也不断涌现。

Web 应用开发的目的是在 Web 的基础设施(包括各种物理设备和软件协议)之上, 采用各种技术来为最终用户提供多种应用服务。因此, Web 应用开发技术在很大程度上决定了 Web 的潜能能否被充分发挥。作为计算机网络技术应用的热点之一, Web 应用开发技术吸引了大量用户的兴趣。从行政、商务到出版、教育等各个行业都对 Web 应用开发提出了迫切的需求。从事 Web 应用开发的人员也不再局限于计算机专业人员, 目前已形成了一个相当大规模的群体。

■ 动机

Web 应用开发涉及的知识面广, 其中涉及大量的名词和概念、协议和规范、方法和技术、语言和工具; 同时, Web 应用开发技术的更新速度快, 新的内容不断涌现。广大从事或者即将从事 Web 应用开发的人员, 希望能够全面、系统、深入、准确地了解 Web 应用开发的基本原理和技术, 并能针对实际的应用需求选择合适的解决方案。

从 2000 年开始, 我们给南京大学计算机系本科生开设一门新课“Web 应用开发技术”, 并编写了“Web 应用开发技术”讲义。该讲义在两次教学实践中逐渐完善, 受到了学生的欢迎, 本书就是在此讲义的基础上编写成的。

■ 编写原则

由于 Web 应用开发涉及的技术内容极为丰富, 因此选择合适的内容是本书编写中的一个难点。本书在内容取舍上遵循以下原则:

- 1) 着重于 Web 应用开发的基本概念、原理和关键技术, 从“应用层交互协议”、“数据表示/展现规范”、“程序设计语言”、“应用开发接口/框架”4 个方面进行讲述。不仅仅局限于其中的某个特定细节, 也不具体介绍各种开发工具、环境的操作和使用方法。

- 2) 于 Web 应用开发中涉及的各种程序开发语言, 包括: C/C++、VBScript 等等, 不罗列语言的语法; 同时也不涉及数据库方面的内容。因为目前已有大量的计算机语言参考手册和有关数据库技术的书籍出版; 而且这些语言以及数据库本身在 Web 应用开发中的使用与

在传统应用程序中并没有本质区别。

3) 对于基于 Java 的 Web 应用开发技术, 本书仅作简要的对比介绍, 不进行详细讨论。因为基于 Java 的 Web 应用开发技术非常丰富, 例如: Sun 公司推出的 J2EE 中包括了 Java Applet、Servlet、JSP、JavaBean 等内容。限于篇幅, 只能待以后专门讨论。

■ 读者对象

作为学习 Web 应用开发的一本参考书。本书适用范围较广, 既可用作高等院校计算机及相关专业的教材, 也可供对 Web 应用开发感兴趣的读者自学, 还可作为相关技术人员的参考书。

本书主要由南京大学计算机科学与技术系王继成、武港山编著。参加本书编著工作的还有孙颖、袁春风、邹飞、苏丰、周源远、郑刚、王强、王庆一等。南京大学计算机科学与技术系张福炎教授审阅了本书中的部分内容, 并提出了许多的意见和建议。此外, 本书在编著过程中还参考了国内外同行编写的大量资料。在此, 对所有为本书提供帮助的人表示衷心的感谢。

由于我们的水平有限, 时间仓促, 书中可能有错误之处; 同时, 由于 Web 应用开发涉及到的知识面广、内容繁多, 书中无法覆盖其中的每个方面, 难免有所疏漏。欢迎广大读者批评指正。我们衷心地希望本书能够给读者带来愉快的学习体验, 并对读者的实际工作起到一定的参考和借鉴作用。

Web 还在不断的发展之中。可以肯定, 各种新的语言、工具, 甚至新的开发技术和方法将不断涌现。但是其基本的原理是类似的。希望通过本书的学习, 读者能够奠定比较好的基础, 在今后的学习和实践中做到触类旁通。

有关此书的程序代码, 请在机械工业出版社网站 www.cmpbook.com 上下载。

编 者

目 录

前言

第一部分 Web 应用开发

第 1 章 Web 应用开发概述	1
1.1 Internet 与 Web	1
1.1.1 Internet 简介	1
1.1.2 Web 简介	3
1.1.3 Internet 技术分类	4
1.2 Web 应用开发	6
1.2.1 Web 发展的需求	6
1.2.2 应用程序发展的需求	8
1.2.3 Web 应用开发的技术内容	10
1.3 本书内容介绍	13
1.3.1 本书的目的	13
1.3.2 内容安排	14
1.3.3 对读者的要求	14
1.4 本章小结	15

第二部分 客户/服务器交互协议

第 2 章 HTTP 概述	17
2.1 HTTP 简介	17
2.1.1 什么是 HTTP	17
2.1.2 HTTP 的功能	17
2.1.3 哪些人需要掌握 HTTP	18
2.1.4 HTTP 中的一些重要术语	19
2.2 HTTP 的特点与历史发展	20
2.2.1 HTTP 的特点	20
2.2.2 HTTP 的历史	22
2.3 与 HTTP 相关的一些规范	23
2.3.1 HTTP 与 MIME	23
2.3.2 HTTP 与 URI	26
2.4 HTTP 的宏观运作原理	27
2.4.1 请求/响应协议	27
2.4.2 客户与服务器通信的三种方式	27
2.5 HTTP 的内部操作过程	30
2.5.1 基于网络层的 HTTP 交互过程	30

2.5.2 一个实际例子	31
2.5.3 使用代理/网关的 HTTP 交互过程	33
2.6 本章小结	35
第 3 章 HTTP 消息	37
3.1 HTTP 消息简介	37
3.1.1 消息类型与格式	37
3.1.2 消息头	38
3.1.3 消息体	39
3.1.4 一个用于查看 HTTP 消息的实用程序	39
3.2 请求消息	41
3.2.1 请求行	41
3.2.2 请求方法	42
3.2.3 请求头	45
3.3 响应消息	51
3.3.1 状态行	52
3.3.2 状态码	52
3.3.3 响应头	56
3.4 实体	58
3.4.1 实体头	58
3.4.2 实体正文	59
3.5 HTTP 消息总结	60
3.6 本章小结	60
第 4 章 HTTP 应用开发与示例	61
4.1 HTTP 应用的分类	61
4.1.1 HTTP 客户程序	61
4.1.2 HTTP 服务器程序	63
4.1.3 服务器端应用程序	64
4.2 HTTP 应用开发方法	64
4.3 基于控件开发 HTTP 客户端的应用	65
4.3.1 Internet Transfer 控件	65
4.3.2 一个 HTTP 客户程序示例	68
4.4 基于 WinInet 开发 HTTP 客户端的应用	72
4.4.1 基本步骤	73
4.4.2 高级特性	78
4.5 基于 Socket 开发 HTTP 应用	83
4.5.1 Socket 简介	83
4.5.2 基于 Socket 开发 HTTP 客户程序	86
4.5.3 基于 Socket 开发 HTTP 服务器程序	88
4.6 HTTP 应用开发总结	95

4.7 本章小结	96
----------------	----

第三部分 数据表示与展现规范

第5章 HTML 基础	97
5.1 HTML 简介	97
5.1.1 超文本与超媒体	97
5.1.2 什么是 HTML	98
5.1.3 HTML 的功能与特性	100
5.1.4 哪些人需要掌握 HTML	101
5.1.5 HTML 的发展历史	101
5.2 SGML 与 HTML	102
5.2.1 SGML 应用与 SGML 文档	102
5.2.2 HTML 中的 SGML 结构类型	102
5.3 HTML 文档的结构	105
5.3.1 文档结构简介	105
5.3.2 文档头	105
5.3.3 文档体	107
5.4 文本	111
5.4.1 空格	111
5.4.2 结构化文本元素	112
5.4.3 行与段落	113
5.4.4 标题	116
5.4.5 列表	116
5.5 表格	119
5.5.1 表格简介	119
5.5.2 表格中的相关元素	119
5.6 本章小结	123
第6章 HTML 提高	125
6.1 超链接	125
6.1.1 超链接与锚简介	125
6.1.2 超链接元素: A 元素	127
6.1.3 文档关系: LINK 元素	128
6.1.4 路径信息: BASE 元素	128
6.2 图像	129
6.2.1 图像: IMG 元素	129
6.2.2 图像映射	130
6.3 可执行对象	133
6.3.1 Java Applet: APPLET 元素	133
6.3.2 通用对象: OBJECT 元素	134

6.3.3 对象的初始化: PARAM 元素	136
6.4 表单	137
6.4.1 表单简介	137
6.4.2 控件元素	140
6.4.3 表单的创建	143
6.4.4 表单的提交	144
6.5 框架	147
6.5.1 FRAMESET 元素	148
6.5.2 FRAME 元素	149
6.5.3 目标框架的指定	150
6.6 样式表	150
6.6.1 样式表简介	150
6.6.2 为 HTML 文档增加样式	153
6.6.3 级联样式表	155
6.7 脚本	157
6.7.1 脚本简介	157
6.7.2 SCRIPT 元素	158
6.7.3 内嵌脚本: 事件属性	160
6.7.4 有关脚本的几点说明	160
6.8 本章小结	161

第四部分 脚本及其在客户端的应用

第 7 章 脚本语言基础	163
7.1 脚本语言简介	163
7.1.1 什么是脚本语言	163
7.1.2 脚本语言的使用	164
7.1.3 脚本语言的种类	165
7.1.4 哪些人需要掌握脚本语言	166
7.2 脚本语言的基本要素	166
7.2.1 数据类型	166
7.2.2 变量与常量	168
7.2.3 运算符	169
7.2.4 流程控制	171
7.2.5 函数与过程	172
7.2.6 其他	174
7.3 脚本语言中对象的使用	175
7.3.1 对象的种类	175
7.3.2 VBScript 中对象的使用	178
7.3.3 JScript 中对象的使用	178

7.3.4 常用对象简介	181
7.4 本章小结	182
第 8 章 脚本在客户端的应用	183
8.1 客户端浏览器与 DHTML	183
8.1.1 客户端浏览器	183
8.1.2 DHTML 与对象模型	184
8.2 浏览器提供的内置对象	186
8.3 文档对象模型与 document 对象	190
8.3.1 文档对象模型	190
8.3.2 元素对象的引用	193
8.4 事件模型与 event 对象	196
8.4.1 事件模型	196
8.4.2 event 对象	199
8.4.3 事件的生命周期	201
8.5 动态样式	202
8.5.1 style 对象	202
8.5.2 样式表对象	203
8.6 本章小结	206

第五部分 Web 服务器端应用开发

第 9 章 服务器端 CGI 程序	207
9.1 CGI 简介	207
9.1.1 网关与 CGI 规范	207
9.1.2 哪些人需要掌握 CGI 应用程序开发	209
9.1.3 CGI 的特点	209
9.2 CGI 的工作原理及其发展	211
9.2.1 CGI 的工作原理	211
9.2.2 WinCGI	213
9.2.3 ISAPI	214
9.3 CGI 规范	215
9.3.1 环境变量	215
9.3.2 标准输入	217
9.3.3 标准输出	219
9.4 CGI 应用程序中表单数据集的处理	222
9.5 服务器端的状态维护	225
9.5.1 服务器端的状态信息	225
9.5.2 返回给客户端的 HTML 文档	227
9.5.3 服务器端存储文件	227
9.5.4 客户端存储文件——Cookie	228

9.6 本章小结	238
第 10 章 服务器端 ASP 脚本程序	239
10.1 ASP 简介	239
10.1.1 什么是 ASP	239
10.1.2 哪些人需要掌握 ASP 脚本程序的开发	240
10.1.3 ASP 的工作原理	240
10.1.4 ASP 脚本程序的特点	242
10.1.5 ASP 与 JSP	243
10.2 ASP 脚本程序的开发	245
10.2.1 脚本语言的使用	245
10.2.2 对象的使用	249
10.2.3 ASP 指令	252
10.3 ASP 脚本宿主中的对象模型	253
10.3.1 一个示例	253
10.3.2 ASP 中的应用程序管理——Application 对象	255
10.3.3 ASP 中的会话管理——Session 对象	258
10.3.4 客户端信息的获取和返回——Request/Response 对象	261
10.3.5 其他内置对象	263
10.3.6 ASP 的实现机制	264
10.4 本章小结	267
第 11 章 专题讨论: Web 信息检索与挖掘	268
11.1 Web 信息检索	268
11.1.1 引言	268
11.1.2 Web 信息检索系统的分类	269
11.1.3 Web 信息检索的关键技术	271
11.1.4 Web 信息检索的结构设计	274
11.1.5 Web 信息检索的性能评价标准	274
11.2 Web 信息挖掘	275
11.2.1 引言	275
11.2.2 Web 信息挖掘的任务	277
11.2.3 文档的特征表示	278
11.2.4 文档摘要	279
11.2.5 文档分类	280
11.2.6 文档聚类	282
参考文献	284

第一部分 Web 应用开发

第 1 章 Web 应用开发概述

学习目标

- 对 Internet 和 Web 以及二者的关系有一个基本的了解。
- 了解互联网技术的三个层次：基础结构、软件开发、应用服务。
- 从 Web 发展和应用程序发展两个角度，理解对 Web 应用开发技术的需求。
- 了解 Web 应用开发的技术内容。

1.1 Internet 与 Web

1.1.1 Internet 简介

(1) Internet 与 Intranet

Internet（互联网）：是一个全球共用的网络，它使用 TCP/IP 协议将遍布全球的许多计算机网络连接起来，从而实现信息资源和服务的共享。

和 Internet 经常一起出现的一个词是 Intranet。所谓的 Intranet，是指一个企业或者组织内部的计算机所组成的网络，该网络采用与 Internet 类似的通信模式，例如 TCP/IP 协议，但仅供企业内部的计算机进行资源和服务共享。因此，Intranet 也可以称为企业网（corporate network）。如图 1-1 所示。为了和 Internet 上的其他计算机进行通信，Intranet 必须通过代理或网关接入到 Internet。在后面的“客户/服务器交互协议”部分，我们将对代理、网关作进一步讨论。

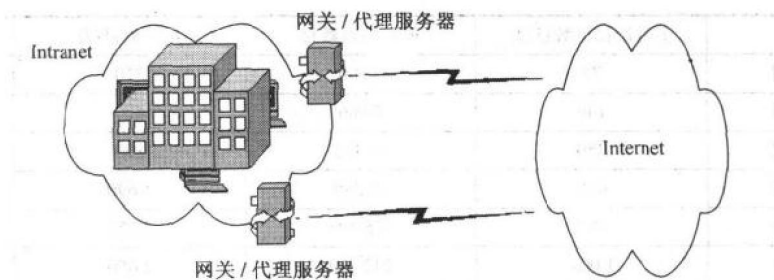


图 1-1 通过代理接入 Internet

(2) Internet 的发展历史

1) ARPAnet

Internet 最初起源于美国国防部高级研究计划局 (Advanced Research Projects Agency, 简记为 ARPA) 在 1969 年建立的网络—ARPAnet。该网络的目的是将研究中心的计算机连接起来以实现相互通信。为此, ARPA 制定了联网计算机之间的通信协议以及路由协议。这些研究工作为 TCP/IP 协议族的制定奠定了基础。

2) NSFnet

随着 ARPAnet 的发展, 越来越多的计算机网络加入到其中, 除了军方以外还有大学、研究所等机构和组织。在 1986 年, 美国国家科学基金会 (National Science Foundation, 简记为 NSF) 建立了网络—NSFnet。之后, NSFnet 就取代了 ARPAnet。

3) Internet

随着网络互联趋势的发展, 出现了 Internet 这样一个词汇, 用以表示全球互联的网络。各个大学、公司、企业、政府机构纷纷建立自己的网络, 并加入到 Internet 之中。目前, Internet 已经构成了全球信息高速公路的基础平台。

(3) Internet 上的服务

正如铁路、公路等交通系统是物质资源流动的动脉一样, 互联网是传输信息资源的动脉。互联网能够传送文本、以及图像、视频等多媒体信息, 提供了 E-mail、Web、FTP 等多种服务。在此基础上, 人们可以方便地实现信息交流、资源共享、网上购物等。因此, Internet 已经成为人们生活中不可缺少的一个部分。

(4) 中国互联网的发展现状

近年来, 我国网络建设也有了很大的发展。中国教育科研网 (China Education and Research Network, 简记为 CERNET)、中国公用网 (China Network, 简记为 CHINANET)、中国科技网 (China Science and Technology Network, 简记为 CSTNET)、金桥网 (China Golden Bridge Network, 简记为 CHINAGBN) 等都具有了相当大的规模, 拥有丰富的中文信息资源和大量的用户。

从 1997 年开始, 中国互联网络信息中心 (China Internet Network Information Center, 简记为 CINIC) 于每年 1 月和 7 月发布《中国互联网络发展状况统计报告》。在报告中, 详细地给出了关于我国 Internet 发展状况的各种统计信息, 包括: 上网计算机数目、Web 站点数目、用户数目及分布、信息流量分布、等等。

表 1-1 统计数据汇总 (1999 年 7 月至 2002 年 1 月)

年 份	上网计算机数目/万	Web 站点数目	用户数目/万	最常用服务
1999 年 1 月	75	5 300	210	电子邮箱 搜索引擎 软件下载
1999 年 7 月	146	9 906	400	
2000 年 1 月	350	15 153	890	
2000 年 7 月	650	27 289	1 690	
2001 年 1 月	892	265 405	2 250	
2001 年 7 月	1 002	242 739	2 650	
2002 年 1 月	1 254	277 100	3 370	
2002 年 7 月	1 613	293 213	4 580	

表 1-1 中汇总了最近三年期间（1999 年 1 月至 2002 年 7 月）历次报告中关于“上网计算机数目”、“Web 站点数目”、“用户数目”、“最常用服务”等状况的统计数据。可以看出，无论是上网计算机数目还是用户数目都以每年 2 倍到 4 倍的速度增长。同时，电子邮件、搜索引擎、软件下载一直是 Internet 上最受人们欢迎的服务。

1.1.2 Web 简介

（1）什么是 Web

万维网（World Wide Web，简记为 Web），是 Internet 上的一种服务，它使用超文本技术将遍布全球的各种信息资源链接起来，以便于用户的浏览。Web 上的信息资源丰富多样，可以是文本、多媒体、数据库，甚至是可执行程序。这些资源可以通过超链接链接起来，在逻辑上形成了一个遍布全球的巨大的“信息网络”。这也是 Web 名称的由来。

我们上网浏览新闻、在线购物、甚至收发 E-mail 都会用到 Web。可以说，Web 已经成为工作和生活中必不可少的一部分。虽然我们平时所说的“上 Internet”在绝大多数情况下是“上 Web”，但 Internet 和 Web 是两个不同的概念。Internet 是 Web 的基础平台，Web 是 Internet 平台上的一种应用层服务，它使得人们能够方便、快捷地发布或获取信息。至于这些信息是如何在 Internet 网络层上传输的，对于 Web 用户而言是透明的。

（2）Web 的组成

Web 有两类主要的组成部分：Web 服务器和 Web 客户端。

1) 服务器（Web 站点）

在物理设备上，是指存放那些供客户访问的信息资源的远程计算机；在软件上，是指能够按照客户的请求将信息资源传送给客户的应用程序。例如：作为南京大学的主页站点在 IP 地址为“202.119.32.12”的一台服务器上，所使用的服务器软件为 Microsoft IIS/5.0。

2) 客户端

在物理设备上，是指客户所使用的本地计算机；在软件上，是指能够接收服务器上的信息资源并展现给客户的应用程序。例如：我们上网时使用的 PC 机就是一个客户端，而浏览 Web 页面的软件通常是 Microsoft Internet Explorer（以下简称 IE）或者是 NetScape Navigator。

（3）Web 的发展历史

Web 最初起源于 20 世纪 80 年代末期的欧洲粒子物理实验室 CERN。当时，Tim Berners-Lee 提出一项计划，其目的是使科学家之间能方便、快速地交流、检索科学资料和数据。为此，Tim Berners-Lee 创建了超文本文档的描述语言，以及在客户和服务器之间传送文档的交互协议，这就是 HTML 和 HTTP 的雏形。

Web 在 CERN 内部得到了广泛的应用，此后逐渐推广到 Internet 上。1993 年，美国伊利诺斯大学（University of Illinois at Urbana-Champaign，简记为 UIUC）的国家超级计算机中心（National Center for Supercomputing Applications，简记为 NCSA）开发了一个带有图形用户界面（Graphics User Interface，简记为 GUI）的 Web 客户端浏览器软件 Mosaic，受到了用户的普遍欢迎。此后，Web 开始迅猛增长。权威杂志《Nature》上发表的研究报告表明，在 1999 年 2 月，Web 上约有 2,800,000 台 Web 服务器，共计 800,000,000 个页面，数据量达到 15TB。而到了 2002 年，仅 Google 所索引的页面数目就已经超过了 20 亿。

目前,无论从用户数目还是从网络流量来看,Web 都是 Internet 上使用最普遍的服务。它已经成为信息发布和信息获取的基础平台,人们可以用它来快速、有效地获取各种信息。随着技术的发展,Web 也逐渐成为互联网应用的基础平台。在此基础上,电子商务、远程教育等新型应用服务也不断涌现。

(4) Web 的支撑技术

在 Web 出现以前,Internet 就提供了 Gopher、FTP 等多种服务。与这些服务相比,Web 提供了前所未有的强大功能,同时,它的使用又非常简捷方便。Web 的成功得益于一系列技术手段的支撑。其中,统一资源标识符(Unified Resource Identifier,简记为 URI)、超文本传输协议(HyperText Transfer Protocol,简记为 HTTP)、超文本标记语言(HyperText Markup Language,简记为 HTML)是三个最为重要和基础的环节。

下面,我们对这些技术作一个简要的介绍,在后面的章节中,我们将会逐一给出详细描述。

1) URI

URI 是对互联网上信息资源进行命名和定位的一种标准机制。在 Internet 上,信息资源可能分布在任何地方,为了让用户能够知道并访问该资源,我们必须采用一种统一的方法为每个资源赋予一个唯一的标识符。该标识符应该包含一些信息以指出如何访问相应的资源,例如:资源所在的服务器名称、资源在服务器上的路径、文件名称、等等。

2) HTTP

HTTP 是在 Web 服务器和客户之间传输信息资源的一种标准协议。为了保证客户和服务端能够彼此理解交互时的语法和语义,二者必须共同遵守一定的交互协议,包括:客户发送请求信息的格式、服务器给出响应消息的格式、等等。例如:客户在下载一个文件时,首先要向服务器提出下载请求,并指定该文件的 URI;而服务器则要向客户报告下载是否成功,并返回对应的资源。

3) HTML

HTML 是在 Web 上采用超文本方式来描述信息资源内容的一种语言规范。为了能够在全球范围内发布信息,我们需要一种能够为所有的计算机所理解的信息资源描述语言。例如:哪些是正文,哪些是标题、图片等。Web 上使用的描述语言是 HTML。你在网上所看到的每一篇文档,无论是新闻报道、天气预报、图书信息等,都是用 HTML 写的。一篇 HTML 文档中还可以包含指向其他文档的链接,这样就在文档之间建立起了关联。

对于用户而言,每次上网浏览的过程都涉及到上述三项技术。例如,当需要浏览新浪网的首页时,首先打开 IE;然后在地址栏中键入新浪网首页的地址“www.sina.com.cn”,在用户键入回车键后,可以在状态栏上看到 IE 正在和新浪网的 Web 服务器进行连接,这使用的就是 HTTP 协议;最后,当网页返回时,IE 按照 HTML 规范对网页进行解析并以可视化的方式展现给用户。如图 1-2 所示。整个过程由客户端软件和服务器软件协作完成,对于最终用户而言是透明的。

1.1.3 Internet 技术分类

Internet 是一系列硬件、软件、协议的集合。在 Internet 的强大功能背后,有着众多技术的支撑。这些软硬件技术成为 Internet 成功的主要因素。随着新技术的不断涌现,Internet 的

功能也会不断地完善和扩充。我们可以将互联网技术分为三个层次，自底向上依次为：基础设施、软件开发、应用服务。



图 1-2 Web 的支撑技术—HTTP、HTML、URI

(1) 基础设施

基础设施是网络的规划、建设人员所关注的部分。从物理设备来看，Internet 的基础设施包括主机、终端、中继设备、以及传输线路等。从协议和软件的角度来看，又包含物理层、链路层、网络层和传输层等。如图 1-3 所示。上述各种物理设备和软件协议构成了 Internet 赖以存在的基础设施。这些技术在“数据通信”、“计算机网络”之类的书籍中会有详细的介绍。感兴趣的读者可以参考相关资料。

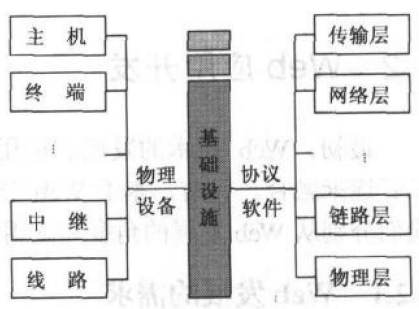


图 1-3 Internet 的基础设施

(2) 应用服务

应用服务是与最终用户直接打交道的部分。从应用服务的目的来看，又可以分为交易服务、信息服务、通信服务、计算服务等。从应用服务的提供手段来看，Internet 的应用服务包括：Web、E-mail、FTP 等多种类型。如图 1-4 所示。这些内容在“电子商务”这类书籍中会有详细的介绍。感兴趣的读者可以参考相关资料。随着 Web 的发展，Web 已经是 Internet 上应用最为广泛的服务、而且它已经逐渐地将 E-mail、FTP 等其他服务融合了起来。因此，在某种程度上，“Internet 应用”已经等同于“Web 应用”。

(3) 软件开发

“软件开发”是“基础结构”与“应用服务”之间的桥梁。Internet 软件开发技术包括各种应用层协议、网络编程技术、Web 应用开发技术等。开发人员的工作目标就是利用这

些技术在 Internet 的基础结构上实现各种应用服务。如图 1-5 所示。本书以及“Java 与 Internet 编程”之类的书籍讲述的都是与 Internet 软件开发相关的内容。

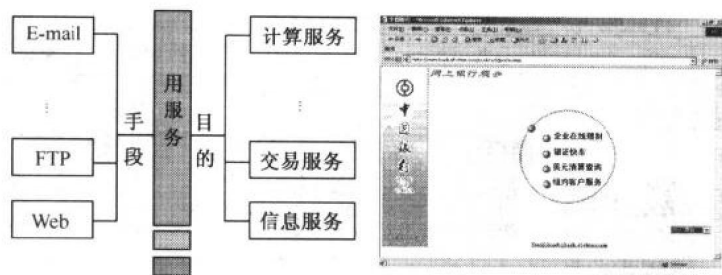


图 1-4 Internet 的应用服务

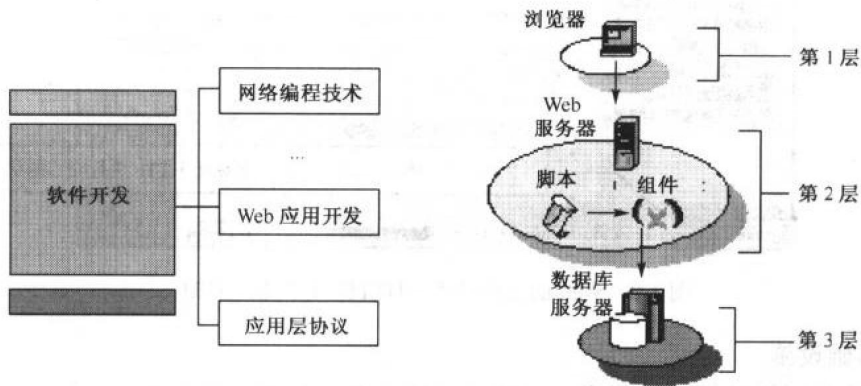


图 1-5 Internet 的软件开发

1.2 Web 应用开发

最初，Web 技术的发展和应用程序开发技术的发展是独立的。在应用需求的驱动下，二者逐渐融合。目前，基于 Web 来开发 Internet 应用已经成为一种主流的解决方案。下面，我们分别从 Web 发展的角度和应用程序发展的角度，来看看对 Web 应用开发技术的需求。

1.2.1 Web 发展的需求

从 1989 年 Web 诞生以来，经过十几年的发展，Web 上的内容已经从静态页面发展为活动页面以及动态页面。

(1) 静态页面

Web 上的内容最早是由一些静态页面组成的。这些页面中可以包含多种媒体元素，丰富了信息资源的表现；页面之间通过超链接彼此关联起来，便于用户浏览、检索。如图 1-6 所示。

静态页面的内容是作者在创建页面时确定下来的，并存储在服务器的文件系统中。除非作者对其进行修改，否则任何用户在任何时候访问它时所得到的内容都是一样的。虽然静态页面已经为用户提供了访问远程信息资源的良好途径，但是却有着很大的局限性。对于信息