

挑战 Java 程序员系列丛书

免费售后培训

免费电子课件及案例源代码

大量论坛技术支持

Java XML

应用程序设计

BitService

北京比特塞威斯图书服务事业部 组编

侯要红 栗松涛 编著



机械工业出版社
CHINA MACHINE PRESS

挑战 Java 程序员系列丛书

Java XML 应用程序设计

北京比特塞威斯图书服务事业部 组编

侯要红 栗松涛 编著



机械工业出版社

本书从程序设计的角度讲解了 XML 的系统知识。本书主要内容包括 XML 基础知识、XML 应用程序、使用 CSS 格式化 XML 页面、使用 XSL 显示 XML 页面、XML 文档类型定义、XML 中的名称空间、XML Schema 概述、XML 的数据源对象、JAXP 与 Java XML 程序设计、Java SAX 程序设计、使用 DOM 管理 XML 文档、使用 JDOM 管理 XML 文档和 XML 对象加载型配置系统设计。本书由浅入深、由易到难,通过丰富的教学案例全面讲解了 XML 的使用。每章都结合具体工程实践,介绍具体项目开发中的应用技术。本书中的程序不需要特殊运行条件和编程环境,有 IE 6、JDK 1.6、JDOM、Xalan 和写字板就可以了。

图书在版编目(CIP)数据

Java XML 应用程序设计 / 侯要红, 栗松涛编著; 北京比特塞威斯图书服务事业部组编. —北京: 机械工业出版社, 2007.8

(挑战 Java 程序员系列丛书)

ISBN 978-7-111-22056-5

I. J… II. ①侯… ②栗… ③北… III. ①Java 语言—程序设计②可扩充语言, XML—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字(2007)第 119948 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

责任编辑: 罗子超

责任印制: 杨 曦

三河市国英印务有限公司印刷

2007 年 9 月第 1 版·第 1 次印刷

184mm×260mm·25.5 印张·630 千字

0001—5000 册

标准书号: ISBN 978-7-111-22056-5

定价: 38.00 元

凡购本书, 如有缺页, 倒页, 脱页, 由本社发行部调换

销售服务热线电话: (010) 68326294

购书热线电话: (010) 88379639 88379641 88379643

编辑热线电话: (010) 88379739

封面无防伪标均为盗版

丛 书 序

在信息技术高速发展的今天，计算机技术已经广泛地应用于各行各业。软件行业作为计算机技术领域的核心产业得到了蓬勃的发展，而软件开发技术作为计算机及其相关行业的关键技术，越来越受到重视。当前，软件开发基本上有两种主流软件开发工具，即 Java 和 .NET。这两种开发工具功能强大、应用领域广泛，是其他开发工具所不能比拟的。可以说，熟练掌握这两种开发工具中的一种，再拥有一定程度的开发经验，就可以成为一名合格的软件开发人员。

目前，相对于大多数职位的人才都面临激烈竞争的情况，合格的软件开发人员却一直是供不应求的。这说明市场上对软件人才需求量很大。但由于软件开发需要丰富的实际工作经验，而对仅仅接受过课本教育，却很少开发实际项目的毕业生来说，是很难立刻胜任软件开发的工作岗位的，从而也就无法达到用人企业的要求，这就是所谓的“人才错位”。而能够有效弥补毕业生与合格的软件开发人员之间差距的途径，目前主要有两种：一是通过上专业培训班或者自己看书学习，并配合相关上机操作自学成才；二是通过用人单位提供实践的机会来获取能力。显然，对刚毕业的学生来说，第一种途径实现起来更容易。

虽然市场上有很多针对 Java 程序设计的参考书和教材，但是，已有的同类书都无法回避的一个事实就是：仅仅依靠纸介质图书和有限的源代码程序来学习 Java 程序设计，不但学习起来非常吃力，而且所得到的结果距离真正的 Java 程序员的需求还有一段距离。我们的结论就是：想真正掌握 Java 这样一种复杂的程序开发工具，除了要学习一套易学易用的图书，还需要看懂大量的实例源代码，以及专家随时的答疑解惑和面对面的学习指导，再加上充足的上机实践。只有这样，才能够真正达到或者接近一个合格的 Java 程序员的水平。

我们开发的这套“挑战 Java 程序员”系列丛书不但图书本身遵循了由浅入深、循序渐进的学习规律，能够使读者轻松快速地掌握 Java 开发的系统知识，而且还配备了完整的售后服务系统，包括：3 个学时的免费培训服务、免费从网上下载 PPT 电子课件、免费下载实例源代码、论坛、E-mail 技术支持服务等。通过这些全方位的售后服务，将 Java 程序员的必备技能依次讲授给读者，为读者铺就一条通向 Java 程序员的成功之路。

本套丛书十分适合于立志成为一个程序员的大中专院校相关专业的学生，或具有一定基础的软件开发爱好者。本套丛书共 6 个品种，包括《Java 面向对象程序设计》、《Java SE 应用程序设计》、《Java XML 应用程序设计》、《Java 数据库应用程序设计》、《Java Web 应用程序设计》和《Java EJB 应用程序设计》，如图 1 所示。

免费售后服务说明

我们为本套丛书的读者提供了丰富的免费售后服务，包括以下 4 方面。

1. 每本书免费赠送 3 个学时的培训（培训地点在北京）

“挑战 Java 程序员”系列丛书中的每本图书自带了 3 个学时的售后培训，购买该书的读者都可以参加。但需要提前预约，否则培训机构不能提供对应的培训内容。

每本图书的售后服务培训班通常根据图书销售情况，周期性推出新班，并通过网站

(<http://www.bits-service.com>) 发布。读者需要提前到网站查询相应班级, 然后通过网络或电话报名预约, 预约电话为 010-64446364 和 010-64446332。预约完的读者在参加培训前需要到网站上或通过电话确认对应班级的变更情况, 培训机构会取消一些人数不多的班, 并将取消信息显示在网站上, 我们将不再一一通知读者, 请读者朋友认真核实。图 2 显示了报名预约的基本流程。

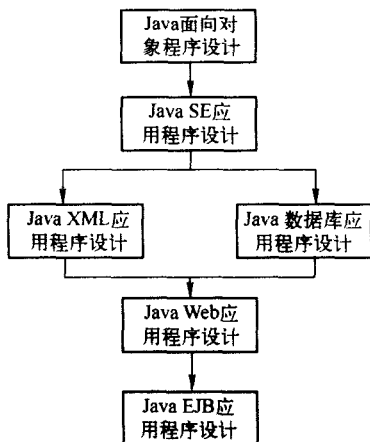


图 1

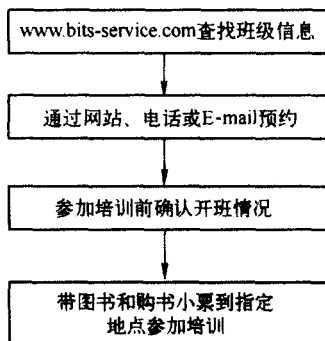


图 2

2. 免费提供精美的 PPT 电子课件

读者可以登录 www.bits-service.com 下载各本图书的电子课件。

3. 免费提供实例源代码

要下载代码的读者, 请首先登录 <http://www.bits-service.com>, 在主页上选择“比特塞威斯图书服务中心 (BITS Book)”, 进入图书服务首页, 在首页菜单条上选择“读者服务”进入“读者服务”频道, 随后可以利用网站的搜索工具或使用标题列表找到对应图书, 在相应的图书网页上单击“源代码下载”超链接, 下载该图书的源代码压缩包。本系列丛书的源代码清单通常按照章节组织, 读者参考图书, 可以使用对应的源代码。

4. 论坛和 E-mail 技术支持

我们为本系列图书提供了专业服务论坛, 网址是 <http://www.bits-service.com> 主页→图书服务论坛→Java 系列图书, 读者在阅读过程中若发现问题, 可以到论坛里发帖子, 培训机构的专业老师会定期进行回复。论坛为每本书提供一个讨论区, 建议读者将问题发送到对应图书的讨论区中, 以便作者及时收集和处理。读者也可以发送 Email 给 booksupport@bits-service.com 寻求专家技术支持。

此外, 我们还会提供一部分特级教师的课堂视频教学文件, 读者可以通过网站下载。

增值服务

除了上述提供的免费服务外, 如果读者希望进一步学习, 我们还提供了一部分增值服务, 包括 Java 项目实习服务、Java 程序员就业推荐服务等。这部分服务需要收取一定费用, 读者如果有兴趣, 可以按照以下联系方式与比特塞威斯公司联系。

增值服务咨询邮箱: allsupport@bits-service.com; 增值服务咨询电话: 010-64446368。

目 录

丛书序

第 1 章 XML 引言	1	2.1.2 使用记事本创建 XML 文档	35
1.1 XML 的发展历史	1	2.1.3 良构的 XML 文档	37
1.1.1 标记语言的出现	1	2.2 XML 文档中的中文编码问题	46
1.1.2 标记语言的特点	2	2.3 XML 的解析器	48
1.1.3 RTF 标记语言	3	2.4 XML 文件的结构	51
1.1.4 HTML 标记语言	7	2.4.1 处理指令	52
1.1.5 标准通用标记语言	11	2.4.2 文件格式定义	52
1.1.6 可扩展标记语言	11	2.4.3 XML 文档的树状结构	53
1.2 XML 基础知识	15	2.4.4 标记的命名规则	54
1.2.1 XML 的定义	15	2.4.5 XML 文档的属性	54
1.2.2 XML 的设计目标	15	2.4.6 XML 文档中的注释	55
1.2.3 XML 的样式语言	16	2.4.7 程序中的文字 CDATA 标记	57
1.2.4 XML 的链接语言	22	2.5 教学信息标记语言设计	59
1.2.5 XML 的名称空间	23	2.6 小结	64
1.2.6 XML 的文档对象模型	23	第 3 章 使用 CSS 格式化 XML	
1.2.7 XML 的文档类型定义	23	页面	65
1.2.8 XML 的文档类型模式	24	3.1 CSS 技术简介	65
1.3 XML 的应用	26	3.2 CSS 文档的基本语法	66
1.3.1 XML 在科学和数学领域中		3.2.1 样式的直接定义法	66
的应用	26	3.2.2 成组选择符	67
1.3.2 XML 在化学领域中的应用	28	3.2.3 利用 CLASS 来定义样式	68
1.3.3 矢量标记语言	29	3.2.4 在 CSS 文件中使用 ID	69
1.3.4 音乐标记语言	30	3.2.5 STYLE 属性的使用	70
1.3.5 资源描述框架	30	3.2.6 CSS 的继承特性	70
1.3.6 人力资源标记语言	30	3.2.7 使用 @import 结合样式表单	71
1.3.7 开放的金融交换标准	31	3.2.8 !important 声明	72
1.3.8 可扩展表单描述语言	31	3.3 将 CSS 和 XML 结合的方法	73
1.3.9 无线标记语言	31	3.3.1 独立样式文件	73
1.3.10 同步多媒体合成语言	32	3.3.2 将样式语句嵌入到 XML 文件	
1.4 小结	33	中	75
第 2 章 XML 应用程序	34	3.3.3 混合方法指定样式	77
2.1 良构的 XML 文档	34	3.3.4 使用多个样式文件	79
2.1.1 问题的描述	34	3.4 CSS 格式化 XML 文档	81
		3.4.1 文字属性	81

3.4.2 定位属性	83	5.2.3 外部 DTD 和内部 DTD 联合使用	164
3.4.3 版式属性	91	5.3 元素声明的语法	165
3.5 在 XML 中使用 HTML 的标记	99	5.3.1 #PCDATA 型数据	165
3.5.1 在 XML 文档中显示图形	99	5.3.2 ANY	166
3.5.2 在 XML 文档中使用超链接	100	5.3.3 声明子标记的一般语法	168
3.5.3 XML 中的换行分段标记	101	5.3.4 可选的子标记	169
3.5.4 在 XML 文档中实现表格	101	5.3.5 复合标记类型定义	171
3.5.5 在 XML 文档中显示表单	102	5.3.6 混合型标记	173
3.6 在 XML 文档中添加脚本语言	104	5.3.7 空标记的定义方法	175
3.7 学生选课管理系统样式设计	105	5.3.8 DTD 中注释的使用语法	175
3.7.1 学生信息样式的设计	106	5.4 实体的声明和使用	177
3.7.2 课程信息的显示	108	5.4.1 内部一般实体	177
3.7.3 教师信息的显示样式	110	5.4.2 外部一般实体	180
3.8 小结	112	5.4.3 内部参数实体	181
第 4 章 使用 XSL 显示 XML		5.4.4 外部参数实体	184
页面	113	5.5 属性的声明和内容类型	185
4.1 XSL 的特点	113	5.5.1 属性的声明	185
4.2 使用 XSL 格式化 XML 页面的例子	114	5.5.2 属性的内容类型	189
4.3 XSL 的模式语言	118	5.6 项目设计(设计 SCLML 学生选课系统标记语言)	197
4.3.1 相对定位和绝对定位	118	5.6.1 设计系统的 DTD 片段	198
4.3.2 XSL 的路径匹配模式	122	5.6.2 设计系统 DTD 文档	199
4.3.3 XSL 的控制语句	137	5.6.3 系统的 XML 数据片段	200
4.4 在 XSL 中使用脚本语句	145	5.6.4 系统的所有数据	201
4.5 多模板 XSL 文件格式	146	5.7 小结	202
4.5.1 多模板 XSL 程序的作用机理	148	第 6 章 XML 中的名称空间	203
4.5.2 多模板程序设计方法	151	6.1 名称空间的基本概念	203
4.6 选课系统 XSLT 样式设计	152	6.2 名称的声明	203
4.6.1 绘制整个系统的 XML 结构图	154	6.3 名称空间的范畴	205
4.6.2 设计学生信息的 XSL 模板	154	6.4 默认的名称	207
4.6.3 设计课程信息的 XSL 模板	156	6.5 使用名称的属性	208
4.6.4 设计教师信息的 XSL 模板	158	6.6 小结	210
4.7 小结	160	第 7 章 XML Schema 概述	211
第 5 章 XML 文档类型定义	161	7.1 XML Schema 的文档结构	211
5.1 文档类型定义	161	7.1.1 XML Schema 文档模板	211
5.2 文档类型声明	161	7.1.2 XML Schema 文档的组成	212
5.2.1 内部 DTD 的声明	161	7.2 标记与属性的定义	213
5.2.2 外部 DTD 的声明	163	7.2.1 简单标记的定义	213
		7.2.2 标记属性的定义	215

7.2.3 标记子标记的定义	217	数据库查询的分页显示	263
7.3 特殊类型属性或标记的定义	218	8.6 小结	264
7.3.1 数值范围的限制方法	219	第9章 JAXP 与 Java XML 程序	
7.3.2 字符串的限制方法	219	设计	265
7.3.3 数值位数的显示方法	222	9.1 Java 处理 XML 概述	265
7.3.4 枚举的显示方法	222	9.2 Java 处理 XML 文件的接口	266
7.3.5 列表类型	223	9.3 Java 常用的解析器	267
7.3.6 联合类型	224	9.4 使用 JAXP 操作 XML 数据	268
7.4 复杂数据类型	224	9.4.1 通过 JAXP 以 DOM 方式操作	
7.5 元素的子元素的出现次数	226	XML	269
7.6 子标记与属性的分组定义	227	9.4.2 通过 JAXP 以 SAX 方式操作	
7.7 Schema 的名称空间	229	XML	271
7.8 XML Schema 中的注释	231	9.5 小结	273
7.9 小结	232	第10章 Java SAX 程序设计	274
第8章 XML 的数据源对象		10.1 SAX 解析基本原理	274
(数据岛)	233	10.2 SAX 解析 XML 的模式	275
8.1 数据岛的一般概念	233	10.2.1 XMLReader 模式	276
8.1.1 DSO 的数据显示方法	233	10.2.2 DefaultHandler 模式	280
8.1.2 内嵌的数据岛	234	10.2.3 HandlerBase 模式	281
8.1.3 外部调用数据的数据岛	236	10.3 SAX 解析属性数据	284
8.2 数据岛的数据显示	238	10.4 使用处理指令处理 XML	
8.2.1 通过简单表格来显示数据岛		数据	286
数据	238	10.5 处理 SAX 解析过程中的错误	
8.2.2 通过嵌套表格来显示数据岛		信息	288
数据	239	10.6 SAX 过滤器	293
8.3 通过记录集管理数据岛	244	10.6.1 SAX 过滤器基本原理	293
8.3.1 创建一个记录集	245	10.6.2 SAX 过滤器程序设计模式	294
8.3.2 通过记录集访问数据岛中各		10.7 小结	297
个标记的方法	247	第11章 使用 DOM 管理 XML	
8.3.3 操作记录集的方法	248	文档	298
8.3.4 遍历记录集	249	11.1 声明一个 DOM 文档对象	298
8.3.5 对记录集进行分页	252	11.2 使用 DOM 技术访问 XML 文档	
8.3.6 把记录取到数组中	255	节点	301
8.4 数据源对象的属性和方法	256	11.3 节点对象的属性	303
8.4.1 数据岛文档的根标记	256	11.3.1 标记属性的读取操作	303
8.4.2 数据岛中的节点	257	11.3.2 标记的类型	305
8.4.3 带有数据类型的 XML 数据岛	259	11.4 节点和属性的创建	307
8.4.4 节点数据类型的访问和设定	262	11.4.1 输出一个 Document 对象	308
8.5 项目应用——使用数据岛实现		11.4.2 创建一个节点	309

11.5 复制某一个子节点	312	12.7.1 JDOM 和 DOM 对象之间的相互 转换	352
11.6 节点的删除与编辑	316	12.7.2 JDOM 和 SAX 对象之间的相互 转换	355
11.7 节点的 XPATH 查询	318	12.8 小结	358
11.8 基于 DOM 的 XSLT 转换	321	第 13 章 XML 对象加载型配置系统 设计	359
11.9 基于 DOM 的合法性验证	324	13.1 XML 对象加载概述	359
11.9.1 基于 DTD 的合法性验证	325	13.1.1 XML 描述对象系统	360
11.9.2 基于 Schema 的合法性验证	327	13.1.2 对象加载系统的设计	362
11.10 小结	329	13.2 Tomcat 对象加载结构	364
第 12 章 使用 JDOM 管理 XML 文档	330	13.2.1 对象加载模块的类结构	365
12.1 JDOM 概述	330	13.2.2 XML 对象加载上下文	365
12.2 XML 文档解析	333	13.2.3 系统的算法模型	367
12.2.1 从文件加载 XML 数据	334	13.2.4 系统的 XML 节点匹配模型	376
12.2.2 从字符串加载 XML 数据	334	13.2.5 系统的规则集合	378
12.3 JDOM 对 XML 的合法性 验证	336	13.3 Tomcat 对象加载行为模型	379
12.3.1 基于 DTD 的合法性验证	336	13.3.1 DocumentHandler 接口所规定的 行为	380
12.3.2 基于 Schema 的合法性验证	338	13.3.2 SaxContext 接口所规定的行为	384
12.4 使用 JDOM 创建 XML 文件	339	13.3.3 算法类工厂方法接口	384
12.4.1 创建一个标记和属性	339	13.3.4 DTDHandler 接口的方法	386
12.4.2 文档类型定义的声明	341	13.3.5 EntityResolver 接口所规定的 方法	386
12.4.3 名称空间的生成	343	13.3.6 XML 文件解析	388
12.4.4 JDOM 节点的复制和移动	345	13.3.7 XMLMapper 对象的初始化	389
12.5 JDOM 节点的导航和枚举	346	13.4 Tomcat 类加载模块的重用	391
12.5.1 JDOM XPath 节点查找	346	13.4.1 使用 Java 程序指定规则	391
12.5.2 JDOM 枚举 XML 数据文件的 节点	347	13.4.2 使用 XML 配置文件制定规则	392
12.6 JDOM 进行 XSLT 转换	349	13.5 小结	393
12.7 JDOM 和 DOM、SAX 接口的 转换	352	附录 软件调试环境	394

第 1 章 XML 引言

本章先讲解 XML 的一些相关概念，如规范的定义、规范的实现、标记语言、专用标记语言、通用标记语言等。尤其是标记语言的概念，希望读者对此要有一定的了解，因为这是学习 XML 语言的基础。随后又重点介绍了 XML 设计目标、XML 样式语言、XML 超链接语言、XML 的名称空间 DOM（文档对象模型）、DSO（数据源对象）和 DTD（文档类型定义）等的基本概念。这些概念在此处只是一个初步的认识。本章的最后还介绍了 XML 在当前的应用情况，以及其发展前景。

1.1 XML 的发展历史

XML 的英文全称是 eXtensible Markup Language（扩展标记语言）。在这里可以简单地理解，XML 就是一种计算机语言。虽然 XML 的应用已经有十多年的历史，但是直到 1998 年 2 月，W3C（全球信息网协会）才正式制定出统一的标准来规范 XML 的使用，这就是我们所说的 XML 1.0 规范。

规范就是计算机业界制定的一种协议，它使某种软件的应用在全世界范围内能达到统一，以方便人们的使用和相互交流。而和规范密切相关的另一个概念是规范的实现。所谓规范的实现就是软件开发商开发合乎规范的软件，但这个实现也需要有计算机厂商的支持。

计算机软件的发展在很大程度上依赖于规范的制定。目前和 XML 相关的许多规范还没有正式通过，但就其目前使用状况而言，该语言的发展方向已经基本确定了。

1.1.1 标记语言的出现

标记语言就是使用某种“记号”来表示特殊信息的语言，如用一种“记号”来表示格式信息或表示数据信息。

格式就是一段文字被显示的样式，如字体的大小、颜色、字型等。CSS 就是专门用来定义 HTML 文档格式的一种语言。

下面是一个具体的例子，如图 1-1 所示。



图 1-1 标记语言示意图

如果上面这段文字要送去排版，那该如何通知排版人员有关文字格式的信息呢？这里使用了在旁边进行说明的办法来解决。这就是一种通俗的标记语言。

如果让计算机理解并按照要求显示上面这段文字，就需要用一种计算机可以“看懂”的语言来解释。具体如下：

我是<bits bold>张三</bits bold>,这本书的<bits italic>作者</bits italic>

上面的文字信息使用了一种标记语言，这种标记语言的详细定义如下。

1. 标记定义

具体的标记定义情况，见表 1-1。

表 1-1 标记定义的作用对照表

标 记 名 字	标 记 作 用
<bits bold></bits bold>	该标记开始和结束之间的内容使用粗体显示
<bits italic></bits italic>	该标记开始和结束之间的内容使用斜体显示

2. 语法定义

如<bits bold>称为 bold 标记的开始标志；前面加一个反斜杠，如</bits bold>，则表示该标记的结束标志。标记的内容是指该标记开始标志和结束标志之间的内容。

这种计算机可以“看懂”的语言，称为标记语言。该标记语言的特点是所有记号都是统一定义的，并且在语法书写上有固定的要求。

1.1.2 标记语言的特点

标记语言是由格式标记（codes）或控制标记（tags）所组成的。这些标记决定了信息的显示格式或数据的意义，但这些标记不能单独存在，只是用来修饰它所包括的文本信息。例如，下面的代码：

我是<bits bold>张三</bits bold>,这本书的<bits italic>作者</bits italic>

<bits bold>、</bits bold>、<bits italic>和</bits italic>是我们所使用的标记，“我是张三，这本书的作者”就是它所要修饰的信息。这些标记本身没有任何意义，但标记结合了文本信息，如“我是张三，这本书的作者”，就表示了一种具有特定格式的文本。

通俗地讲，标记语言就是有关使用什么样的记号和如何使用这些记号的一种语言。通过它，可以和计算机进行自由交流。像所有其他计算机语言一样，它有一定的词汇，但这些词汇数量不多。如有名的 HTML 标记语言，总词汇（标记）量也不超过 100 个。除此之外，

它还有一定语法, 这些语法用来控制如何使用词汇。

目前, 标记语言的种类很多, 根据具体的使用范围可以分为“专用标记语言”和“通用标记语言”两种。

1. 专用标记语言

专用标记语言用来表示某种特殊的信息, 通常被某一种或几种应用软件所支持, 是因特殊用途的需要而发展起来的。例如, HTML 语言就是一种有特殊用途的标记语言。该标记语言是为了因特网上的网页而专门设计的, 其重点在于信息的显示, 是 IE 或 Netscape 等浏览器都普遍支持的一种语言。

例如, Word、WPS、写字板等应用软件, 也都是通过标记语言中的标记信息来处理文本中所要求的特殊格式。它们所支持的标记语言也属于专用标记语言的范畴。

2. 通用标记语言

通用标记语言的设计并不针对某一特殊应用软件, 或某一特殊用途, 通常而言, 它只描述文件中文字的内容与结构, 是一般化的语言。不太严格地讲, 它还是一种定义语言的语言。标记语言本身的一般化, 使许多应用软件能够阅读它, 从而提供了一种在不同的应用软件之间交互交流数据的平台, 如图 1-2 所示。这种语言的特点是具有被国际上的标准化组织通过了的并被全世界普遍接受的标准。例如, SGML (标准通用标记语言) 就是一种通用标记语言 (SGML 也是一种国际标准)。使用该标记语言, 一方面便于计算机的处理, 另一方面便于不同的计算机, 甚至不同的应用软件之间的数据共享。

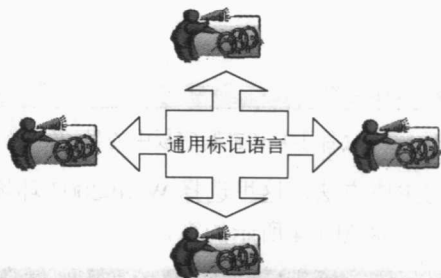


图 1-2 通用标记语言示意图

1.1.3 RTF 标记语言

RTF (Rich Text Format) 是在字处理软件中广泛应用的一种标记语言, Word 和写字板都支持它。和其他的标记语言一样, RTF 语言自身定义了许多标记, 这些标记可以表示字体格式的信息。下面通过例子来体会这种语言的特点。具体见程序 1-1。

程序 1-1

文件 1-1.rtf

```
{\rtf1\ansi\ansicpg936\deff0\deflang1033\deflangfe2052
  {\fonttbl
    {\f0\fswiss\fcharset0 Arial;}
    {\f1\fnil\fcharset134 \cb\ce\cc\ce5;}
  }
```

```
\viewkind4\uc1\pard\b\f0\fs40  
hello the world  
\lang2052\b0\fs20\par  
}
```

通过程序 1-1 可以发现, 在多行程序中只有一行是关键性的信息 (hello the world), 其他都是有关格式的控制信息, 即用来表示文字的格式显示。图 1-3 是这段程序在写字板中的显示结果。

通常不会要求通过手工输入这些标记的方法建立 rtf 文件, 而是直接在一些支持 rtf 格式的工具软件中自动生成。

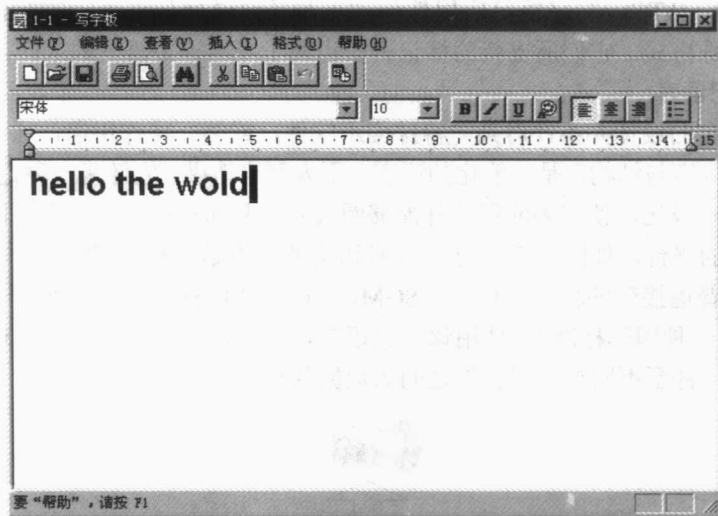


图 1-3 文件 1-1.rtf 在写字板中的显示结果

下面举例说明这种文件的生成方法, 这里选择 Word 2003 环境进行讲解。

(1) 打开 Word 2003, 进入如图 1-4 所示的窗口。

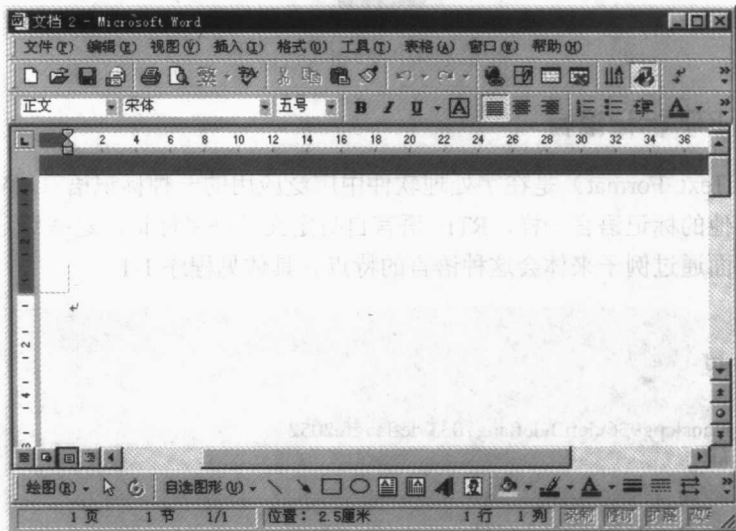


图 1-4 Word 2003 窗口

(2) 创建文档。在文件中输入“hello the world”，将单词“hello”设成红色，其最终结果如图 1-5 所示。

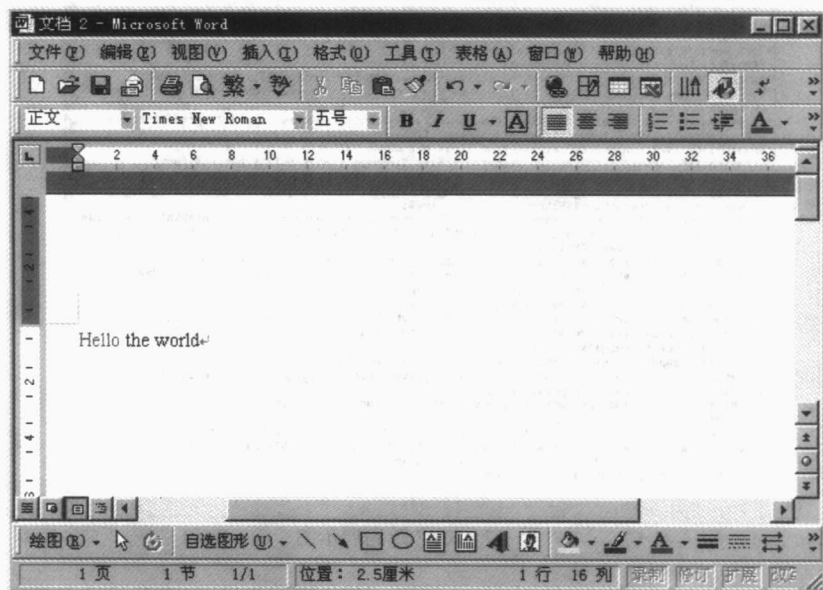


图 1-5 第二步结果图

(3) 将该文档保存成 RTF 文件格式，保存格式如图 1-6 所示。



图 1-6 保存文件

该文件已保存到“My Documents”文件夹中。使用记事本程序打开的结果，如图 1-7 所示。

Word 2003 只是这些代码的生成工具，是一种代码生成器。读者也可以在记事本中手工

创建该文件（但是要足够细心），其效果是一样的。

第 1 章



图 1-7 RTF 源码展示图

严格地讲，Word 文档也是标记语言生成的，但当使用记事本打开一个*.doc 文件时，得到的结果却与 RTF 文件有很大不同，如图 1-8 所示。

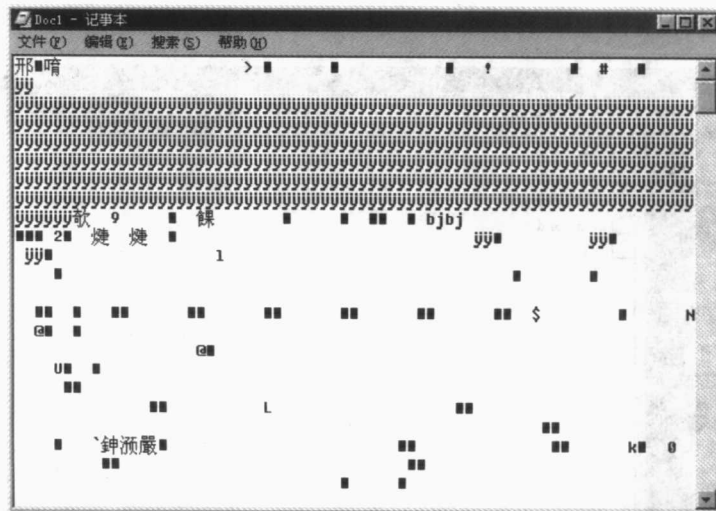


图 1-8 Word 文档在写字板中打开的结果

为什么会出现这种情况呢？这是因为 Word 应用软件在生成标记语言文件后又将该文件转变成了二进制的形式。

上面介绍了 RTF 标记语言的一般概念，这种标记语言在字处理软件领域有着十分广泛的应用。下面介绍另一种标记语言，这种标记语言在 Internet 上被广泛应用。

1.1.4 HTML 标记语言

HTML (HyperText Markup Language, 超文本标记语言) 是一种专门用于格式化 Web 数据的语言。该语言简单明了, 在 Web 技术的推广方面起着很大的作用。下面通过一个例子来认识一下这种标记语言的特点。具体见程序 1-2。

程序 1-2

文件 1-2.htm

```
<html>
<head>
<title> hello the world </title>
</head>
<body>
<h1 align="center">hello the world</h1>
</body>
</html>
```

该文件的作用很简单, 只在 IE 浏览器的窗口中显示 “hello the world”, 在 IE 6.0 中的显示结果, 如图 1-9 所示。

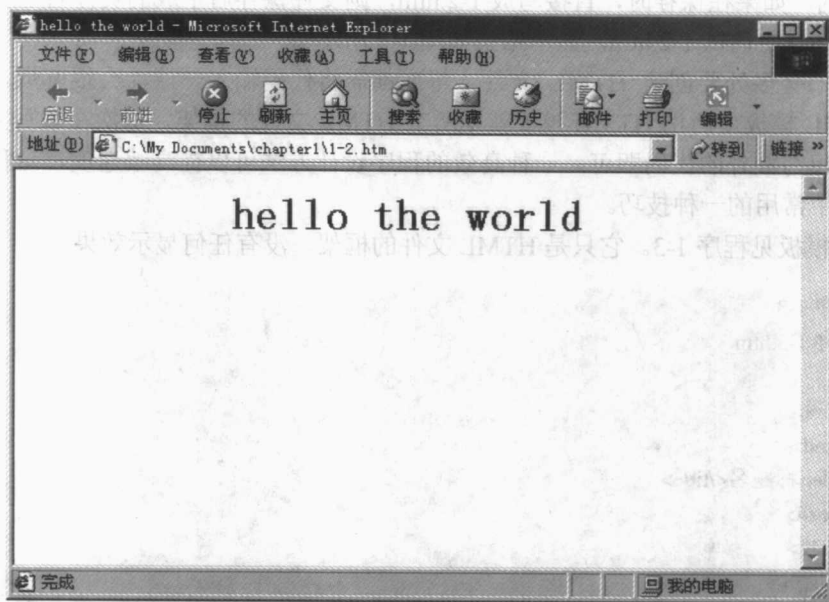


图 1-9 HTML 标记语言在 IE 6.0 中的显示结果

这里强调的是, 本书几乎所有的程序都是在记事本中创建的, 调试工具只使用 IE 6.0 和记事本。下面的例子显示了如何调试程序 1-2。

- (1) 打开记事本。
- (2) 在该记事本中输入程序 1-2 的代码。然后保存, 保存的设置如图 1-10 所示。

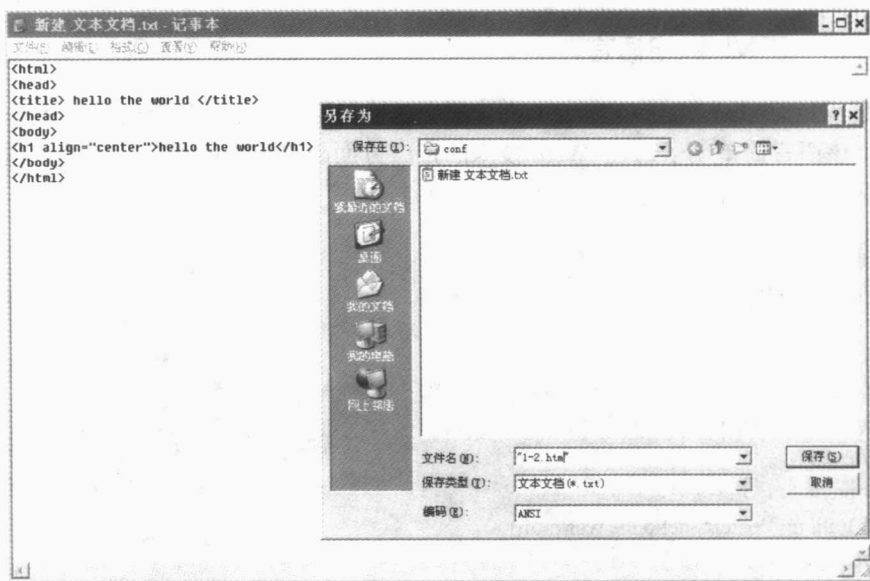


图 1-10 文件的保存方法

在“文件名”文本框中输入“1-2.htm”，即强制性的将文件命名为 1-2.htm，表示它的格式是 HTML 的。如果在保存时，直接写成 1-2.htm，则文件保存后生成的文件名是 1-2.htm.txt。保存后从 IE 6.0 中调试 1-2.htm，显示结果如图 1-9 所示。

由于本书有多处要创建 HTML 文件，每次都需将代码输入一遍，会很麻烦，下面将建立一个 HTML 模板。所谓 HTML 模板就是所有 HTML 文件的框架，每次在编写文件时，只要在框架内填写额外的代码即可。一种高效的程序设计方法可以提高编速度，这种方法是程序开发过程中常用的一种技巧。

HTML 模板见程序 1-3。它只是 HTML 文件的框架，没有任何显示效果。

程序 1-3
文件 1-3.htm

```
<html>
<head>
<title>标题名</title>
</head>
<body>
添加代码
</body>
</html>
```

1. HTML 标记语言的优点

(1) 标记有限。HTML 的所有标记数量不超过 100 个，而常用的标记就只有几十个。这些标记是国际上公认的通用标记，用户只要记住每一个标记的用法，就可以在 Internet 上发布网页。正因如此，HTML 一出现就在全世界得到了很快的普及。