



本书内附光盘

网络精英系列

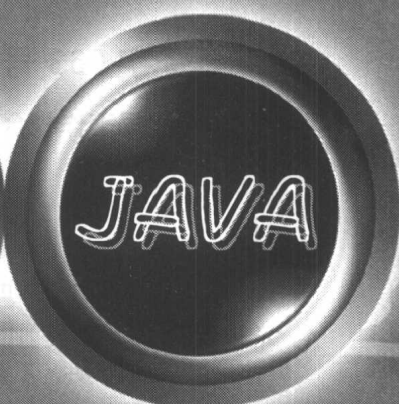
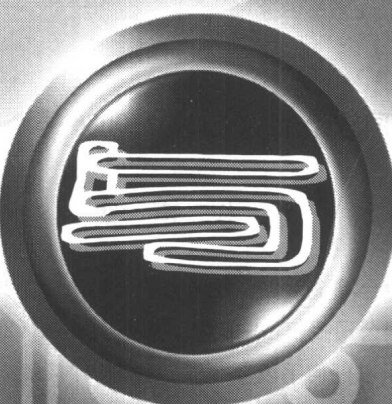
XML工作室 陈锦辉 王景皓 编著

XML与JAVA 程序设计大全

本书特色

本书重点放在XML与Java的程序设计以及相关XML分析器的解说(如: Xerces、SAX、DOM等)。编排上采用交错引导的方法解说。首先说明什么是XML以及XML与Java的相互关系。从第3章开始,正式进入相关技术的详细说明,基本上单数(3、5、7)章为XML的相关技术说明,而双数(4、6、8)章则为使用Java实际编写XML的应用程序,第9章则将重点放在说明JDK对XML的相关支持,第10章为数据库的集成应用,最后一章中,我们以一个实际的范例通过Cocoon来实际编写一个Java/XML的应用程序,将XML文件转变为各种多样化的显示方式。书中光盘内附有范例程序源代码,此书是一本理论与实战相结合的最佳“XML + Java=Portable”入门与进阶的好书,阅后相信您一定会感到物超所值!

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE



程序设计大全

XML 工作室 陈锦辉 王景皓 编著

中国铁道出版社

2002年·北京

(京)新登字 063 号

北京市版权局著作权合同登记号: 01-2002-0086 号

版 权 声 明

本书中文繁体字版由台湾金禾资讯股份有限公司出版(2001)。本书中文简体字版经台湾金禾资讯股份有限公司授权由中国铁道出版社出版(2002)。任何单位或个人未经出版者书面允许,不得以任何手段复制或抄袭本书内容。

图书在版编目(CIP)数据

XML 与 JAVA 程序设计大全/陈锦辉, 王景皓编著. —北京: 中国铁道出版社, 2002. 1

ISBN 7-113-04510-3

I. X… II. ①陈… ②王… III. ①可扩充语言, XML—程序设计

②JAVA 语言—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字(2001)第 097444 号

书 名: XML 与 JAVA 程序设计大全

作 者 XML 工作室 陈锦辉 王景皓

出版发行: 中国铁道出版社(100054, 北京市宣武区右安门西街 8 号)

策划编辑: 严晓舟

特邀编辑: 王占清

封面设计: 孙天昭

印 刷: 北京市彩桥印刷厂

开 本: 787×1092 1/16 印张: 28 字数: 677 千

版 本: 2002 年 2 月第 1 版 2002 年 2 月第 1 次印刷

印 数: 1~5000 册

书 号: ISBN 7-113-04510-3/TP·666

定 价: 46.00 元

版权所有 盗版必究

凡购买铁道版的图书, 如有缺页、倒页、脱页者, 请与本社计算机图书批销部调换。

出版说明

XML 是一种具有描述数据功能的语言, 因此十分适合作为知识表示语言或作为描述组件及文件格式的表示方法。它可以让数据在不同的来源中, 根据通用的语法规则来处理。而 Java 则提供了一个跨平台的方案, 再加上 Java 本身就出生于互联网, 对于分布式处理环境的支持更是不在话下, 这也是为什么大家都说 “XML + Java=Portable”, 因为 XML 是一种数据的表示法, 而 Java 则是一种程序设计的方法, 所以两者缺一不可。

本书所有范例在随书所附光盘 “Sample” 文件夹下。

若您对本书的程序问题有疑问, 请写 E-Mail: xml@eknow.com.tw

本书由台湾金禾资讯股份有限公司提供版权, 经中国铁道出版社计算机图书项目中心审选, 张翰文、李自远、葛兰、彭涛、马超、杨军、陈贤淑及廖康良等同志参与了本书的整稿及编排工作。

2002 年 2 月

目 录

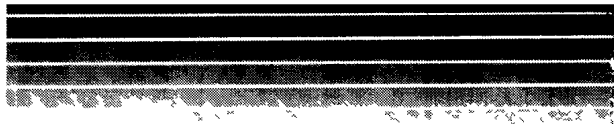
第 1 章 本书导读与随书光盘介绍.....	1
1-1 本书背景与目的.....	2
1-2 本书简介.....	2
1-3 本书范例.....	2
1-4 浏览器的选择.....	3
1-5 本书练习题.....	3
1-6 本书光盘使用说明.....	3
1-7 系列丛书说明.....	3
第 2 章 XML 与 Java.....	5
2-1 XML 简介.....	6
2-1-1 SGML、XML 与 HTML.....	6
2-1-2 XHTML.....	8
2-1-3 命名空间 (Namespace).....	14
2-1-4 XSL.....	14
2-1-5 验证规则 (DTD 与 XML Schema).....	15
2-1-6 XLink.....	16
2-1-7 XPath 与 XPointer.....	16
2-2 XML 应用.....	16
2-2-1 传统数据处理.....	17
2-2-2 文件驱动程序设计方式.....	17
2-2-3 软件组件状态保存.....	17
2-2-4 企业数据绑定.....	18
2-3 Java 简介.....	19
2-4 Java 和 XML 的结合.....	19
2-4-1 应用层.....	20
2-4-2 表示层.....	20
2-4-3 核心层.....	22
2-4-4 链路层.....	22
2-4-5 资源层.....	22
2-5 总结.....	23
练习.....	23
第 3 章 Well-Formed XML.....	25
3-1 建立第一份 Well-Formed XML 文件.....	26

3-1-1	Well-Formed XML 分析器.....	26
3-1-2	XML 的基本语法.....	26
3-1-3	xml 与 version 声明.....	27
3-1-4	根结点 (Root).....	30
3-1-5	所有的标记必须以树状排列.....	32
3-1-6	成对的标记出现方式.....	35
3-1-7	空的标记结尾必须加上"/".....	37
3-1-8	标记名称与属性的规定.....	39
3-1-9	属性值前后必须被“ ” 所包围.....	41
3-1-10	特殊字符的规定.....	43
3-2	设计 XML 文件时必须的考虑.....	45
3-2-1	标记的考虑.....	45
3-2-2	属性的考虑.....	48
3-3	语法.....	49
3-3-1	用 XML 设计的标准语言.....	49
3-3-2	建立属于您的语言语法.....	50
3-4	使用命名空间避免冲突.....	50
3-4-1	XML 为何需要命名空间.....	50
3-4-2	命名空间如何解决问题.....	52
3-4-3	在 XML 中如何使用命名空间.....	53
3-4-4	XML 使用命名空间的典型——XSL 与 XHTML.....	55
练习		58
第 4 章	分析 XML 文件.....	61
4-1	准备工作.....	62
4-1-1	分析器的获取.....	62
4-1-2	SAX 相关包的获取.....	63
4-1-3	准备好一份 XML 文件.....	65
4-2	回调函数.....	66
4-3	读取 XML.....	68
4-3-1	初始化设置.....	68
4-3-2	分析文件.....	70
4-4	内容处理器 (Content Handler).....	72
4-4-1	文件定位器 (Document Locator).....	78
4-4-2	文件的开始与结束.....	79
4-4-3	处理命令.....	80
4-4-4	命名空间.....	81
4-4-5	元素.....	84
4-4-6	元素数据.....	87

4-4-7	空格	90
4-4-8	被忽略的对象	91
4-4-9	设置分析器的内容处理器.....	91
4-5	错误处理器.....	93
4-5-1	警告	96
4-5-2	可恢复的错误	97
4-5-3	不可恢复的错误.....	98
4-5-4	产生错误	99
4-6	加载分析器的正确方法.....	100
练习		105
第 5 章	验证规则 —— DTD 与 XML Schema	107
5-1	ValidatedXML 文件	108
5-1-1	DTD	108
5-1-2	使用 DTD 的优缺点	109
5-1-3	Validated XML 文件.....	110
5-1-4	如何使用 Xerces 做验证工具	110
5-2	内部 DTD 声明与外部 DTD 声明.....	111
5-2-1	内部的 DTD 声明	111
5-2-2	外部的 DTD 声明	114
5-2-3	内部的 DTD 与外部 DTD 的合作	118
5-3	元素 (ELEMENT)	124
5-3-1	非空元素的声明与使用.....	124
5-3-2	空元素的声明与使用 (EMPTY 的使用)	127
5-3-3	不做限制的元素声明与使用 (ANY 的使用)	133
5-4	树状结构.....	135
5-5	结构符号.....	136
5-5-1	“()”、“,” 结构符号与没有结构符号	137
5-5-2	重复性符号 “*”、“+”	140
5-5-3	选择性符号 “?”、“ ”	145
5-6	更复杂的文件结构与 DTD 声明.....	150
5-6-1	混合型的声明	150
5-6-2	递归式的声明	151
5-7	实体.....	153
5-7-1	内部实体与外部实体.....	154
5-7-2	可拆解实体与不可拆解实体.....	161
5-7-3	实体的种类	164
5-8	再谈实体.....	164
5-8-1	在 DTD 声明中使用实体引用	164

5-8-2	参数型的实体引用.....	169
5-8-3	一般型与参数型实体引用的合作.....	172
5-8-4	多个外部 DTD 的合作方式	174
5-9	其他命令.....	177
5-9-1	备注	177
5-9-2	IGNORE 与 INCLUDE	178
5-9-3	CDATA 关键字	180
5-10	元素的属性.....	182
5-10-1	属性使用时的 Well-Formed 限制.....	182
5-10-2	属性的声明语法.....	182
5-10-3	属性(值)类型.....	183
5-10-4	属性的内建值类型.....	183
5-11	属性值类型.....	184
5-11-1	CDATA	184
5-11-2	Enumerated.....	186
5-11-3	NMTOKEN.....	188
5-11-4	NMTOKENS	190
5-11-5	ENTITY	191
5-11-6	ENTITIES.....	197
5-11-7	NOTATION.....	200
5-11-8	ID	201
5-11-9	IDREF.....	204
5-11-10	IDREFS.....	208
5-12	描述属性内建值类型.....	210
5-12-1	#IMPLIED	210
5-12-2	#REQUIRED	211
5-12-3	#FIXED.....	214
5-12-4	特定的默认字符串.....	219
5-13	内建属性名称.....	223
5-13-1	xml:space 内建属性.....	224
5-13-2	xml:lang 内建属性.....	227
5-14	XML Schema 简介.....	229
5-14-1	XML Schema 在 W3C 的审议进度.....	229
5-14-2	XML Schema 与 DTD	230
5-15	XML Schema 基础.....	231
5-15-1	XML Schema 的基本语法.....	231
5-15-2	W3C 与 Microsoft 的 XML Schema	232
5-15-3	指定 XML Schema.....	233
5-16	XML Schema 的结构(元素)声明	234

5-16-1 XML 元素声明<ElementType>.....	234
5-16-2 XML 子元素声明<element>.....	236
5-16-3 组元素<group>.....	239
5-16-4 属性声明<AttributeType>.....	241
5-16-5 元素属性声明<attribute>.....	243
5-17 XML Schema 的数据类型声明.....	246
5-17-1 数据类型分类.....	247
5-17-2 数据类型声明（直接声明）.....	248
5-17-3 数据类型声明（使用<datatype>声明）.....	250
练习.....	252
第 6 章 验证 XML 文件	255
6-1 设置分析器.....	256
6-1-1 设置属性和特征.....	256
6-1-2 激活验证功能.....	257
6-2 错误处理器.....	261
6-2-1 警告(Warning).....	263
6-2-2 错误(Error).....	263
6-2-3 重大错误(Fatal Error).....	265
6-3 DTD 处理器.....	268
练习.....	272
第 7 章 XSL	275
7-1 通过 XSL 转换 XML 文件.....	276
7-1-1 XSL 处理器 (XSL Processor).....	277
7-1-2 转换与格式化排版.....	277
7-1-3 XSL 的主要概念.....	279
7-1-4 XML/XSL.....	279
7-2 如何使用 XSL 排版样式表.....	280
7-3 XPath 路径表示法.....	282
7-3-1 XPath 简介.....	282
7-4 模板 (Template) 的使用方法.....	284
7-4-1 xsl:stylesheet 元素.....	284
7-4-2 模板 (xsl:template 元素) 匹配原则.....	286
7-4-3 xsl:for-each 元素.....	290
7-4-4 xsl:apply-templates 元素 (应用样式).....	293
7-4-5 xsl:value-of 元素 (取出 XML 元素或属性数据).....	297
7-5 在输出中增加 XML 数据.....	300
7-5-1 xsl:comment 元素 (加入注释).....	300
7-5-2 xsl:copy 元素 (复制数据).....	303



7-5-3	xsl:attribute 元素 (增加属性)	304
7-5-4	xsl:element 元素 (增加标记)	307
7-6	XSL 的条件语句控制功能	311
7-6-1	xsl:if 元素	311
7-6-2	xsl:choose、xsl:when、xsl:otherwise 元素	315
练习		320
第 8 章	使用 DOM	323
8-1	什么是 DOM	324
8-1-1	DOM 与 SAX 的比较	326
8-1-2	DOM 和 Java 的关系	327
8-2	DOM 分析器	327
8-2-1	建立 DOM 分析器	327
8-2-2	取得 DOM 文件	329
8-2-3	使用标准方法取得 DOM 文件	330
8-2-4	一个更好的输出	331
8-3	DOM 的类	333
8-3-1	MVC 设计方式	335
8-3-2	DOM 基本操作	336
8-3-3	继承式和扁平式	340
8-3-4	结点类型	341
8-3-5	文字 (Text)	343
8-3-6	元素 (Element)	344
8-3-7	属性 (Attribute)	345
8-3-8	注释 (Comment)	346
8-3-9	文件 (Document)	347
8-3-10	DTD (Document Type)	348
8-3-11	处理命令 (ProcessingInstruction)	349
8-3-12	实体引用 (Entity Reference)	350
8-3-13	字符数据 (CDATA)	351
8-4	进行验证	352
8-5	纵横 DOM	353
8-5-1	使用 NodeIterator	354
8-5-2	使用 TreeWalker	358
8-5-3	使用 NodeFilter	359
练习		362
第 9 章	JDK 对 XML 的支持	363
9-1	简介	364
9-2	JAXP	365

9-2-1 DOM.....	365
9-2-2 SAX	368
9-2-3 XSLT.....	372
9-3 JAXB.....	375
9-4 JAXM	377
练习	378
第 10 章 XML 与数据库的应用	379
10-1 数据库.....	380
10-2 JDBC 基础	382
练习	391
第 11 章 XML 的多样化显示.....	393
11-1 全新的网站形式.....	394
11-2 配置一个 Cocoon 的环境.....	394
11-2-1 安装 Java	395
11-2-2 安装 Tomcat	395
11-2-3 安装 Cocoon.....	397
11-3 初探动态 XML 文件.....	402
11-4 多样的动态内容.....	403
11-5 使用 servlet 来建立动态 XML 文件	407
11-6 Cocoon 的未来	412
练习	413
附录 A XML 的 EBNF 语法.....	415
XML 的 EBNF 语法	416
附录 B XML 分析器的特征(Features).....	421
一般特征	422
DOM 特征	424
SAX 特征.....	425
附录 C XML 分析器的属性(Properties)	427
一般属性	428
DOM 分析器属性	428
SAX 分析器属性.....	428
附录 D ISO-639	431
ISO-639.....	432

本书导读与随书光盘介绍

第 1 章

1-1 本书背景与目的

在本书中,我们将重点介绍 Java 与 XML 的程序设计以及相关 XML 编译器(如: Xerces、SAX、DOM 等)。

XML 提供了可以交换的信息模式,它可以让数据在不同的来源中,依据统一的方法来处理。而 Java 则提供了一个跨平台的方案,再加上 Java 就是出身于因特网,所以对于分布式处理环境的支持又是不在话下。这也就是为什么大家都说 Java+XML=Portable,因为 XML 是一种数据的表示方法,而 Java 则是一种程序设计的方法,就像一首好歌一定要给好的歌手唱才会好听,缺一不可。

1-2 本书简介

本书共分为 11 章,在第 1 章中我们将介绍本书的背景与目的,并提供阅读本书所需要的程序安装方法(随书光盘)及系统需求。

剩余 10 章,我们介绍了 XML 以及 Java 对它的支持。在第 2 章中,我们将首先说明什么是 XML 以及 Java 与 XML 的关系。从第 3 章开始,正式进入相关技术的详细说明。基本上双数(4、6、8)章为 XML 的相关技术说明,而单数(5、7、9)章则使用 Java 实际编写 XML 的应用程序。

第 9 章则将重点放在说明 JDK 对 XML 的相关支持,第 10 章为数据库的集成应用。最后一章中,我们以一个实际的范例通过 Cocoon 来实际编写一个 Java/XML 的应用程序,将 XML 文件转变为各种多样化的显示方式。

以下是本书章节简介

- 第 1 章 本书导览与随书光盘介绍
- 第 2 章 XML 与 Java
- 第 3 章 Well-Formed XML
- 第 4 章 分析 XML 文件
- 第 5 章 验证规则——DTD 与 XML Schema
- 第 6 章 验证 XML 文件
- 第 7 章 XSL
- 第 8 章 使用 DOM
- 第 9 章 JDK 对 XML 的支持
- 第 10 章 XML 与数据库的应用
- 第 11 章 XML 的多样化显示

1-3 本书范例

本书的范例有正确与错误之分,识别以下面规则为主:

- (1) 凡是前无说明该范例是正确或错误范例标题者,一律为正确范例。
- (2) 凡是前有正确范例标题者,都为正确范例。

(3) 凡是前有错误范例标题者，都为错误范例。

本书范例中的行号都是为了说明方便起见所设置的，所有的实际文件内容都不应该含有这些行号。



1-4 浏览器的选择

因此，本书有些章节必须使用浏览器，我们采用的是 Microsoft Internet Explorer 5.x，读者在阅读本书前应该安装 IE 5.x 并且设置为默认的浏览器以方便执行范例。

1-5 本书练习题

本书各章（不含第 1 章），每章后面都附有练习题，作为读者学习完该章后复习所学来用。

1-6 本书光盘使用说明

随书光盘的内容：

“Sample”目录：内含本书所有的范例。在第 5 章中还包含有我们所打包的 MySAXParser 分析器。

1-7 系列丛书说明

本书“XML 与 Java 程序设计大全”是“网络精华录系列”所推出的又一本书，我们将目标锁定为已经熟悉 Java 语法但对于 XML 还很陌生，却希望完成 Java/XML 应用程序的读者。因此，在本书中，我们仍然使用的 3 个章节来说明 XML 的相关基本知识，包含有“Well-Formed XML”（第 3 章）、“XML 验证规则”（第 5 章）、“XSL”（第 7 章）。不可避免的，这 3 章的内容取材自系列丛书《XML 与 ASP 网站实作大全》，并做适当的修改（例如：分析器由 MSXML 改为 Xerces），而这 2 本书的目标读者群分别为使用 Java 与 ASP 的程序员，在此特别提出说明。

XML 与 Java

第 2 章

在本章中,将要介绍 XML 基本概念和相关的一些技术,让您能够有一个全面的了解。同时我们也会介绍 Java 对 XML 有哪些支持和应用。

2-1 XML 简介

XML (eXtensible Markup Language) 是一种具有数据描述功能 (Meta-language)、高度结构性及可验证性的语言。和 HTML 一样,XML 同样使用了标记与属性;而和 HTML 最大的不同点则在于 XML 的标记与属性允许用户自行定义,并可以依照所定义的标记与属性的语法来开发应用程序。在 XML 文件中,可以使用标记来描述数据,或配合属性来辅助描述数据,因此,XML 十分适合用于作为对象或标准的描述语言;并且由于可以借助验证规则 (DTD 或 XML Schema) 来规范一个 XML 文件的内容与结构,所以 XML 又很适合用于做 B2B 或 B2C 的数据交换格式。在本节中,将简单介绍许多与 XML 相关的技术,其中包含有 XHTML、XSL、验证规则 (DTD 与 XML Schema)、命名空间、XLink、XPath、XPointer 等。

2-1-1 SGML、XML 与 HTML

本小节中,将说明 SGML、XML、HTML 与 XHTML 四种标记语言。首先,HTML (Hyper Text Markup Language) 是大家最熟悉的标记语言了,它也是目前最流行的网页设计语言,但是,许多人却不知道 HTML 实际上是由另一个语言所制定出来的。而制定 HTML 语言的便是 SGML。

SGML

SGML (Standard Generalized Markup Language) 为国际标准化组织 (ISO) 于 1986 年所通过认证的一个标准通用标记语言。它有别于特殊用途的标记语言;也就是说,SGML 并不是针对某一特定应用软件所设计的,而是为了描述文件中文字的结构与所代表的意义而设计的。以下是 SGML 的特性:

SGML 的特性:

- (1) 描述语言 (Meta Language): 所谓描述语言,就是一个语言可以用来描述或产生另一种语言,因此描述语言必须拥有可扩充性,SGML 就是一种描述语言,例如: HTML 就是由它所产生的语言。换句话说,当初制定 HTML 的那一群人,是利用 SGML 制定出一个适用于 WWW (World Wide Web) 上使用的标记语言。
- (2) 标记语言 (Markup Language): SGML 也是一个标记语言,即 SGML 也是用一些字符来代表文字的涵义,这些字符在 SGML 中是以标记的形式出现 (类似于 HTML),所以 SGML 也可称为一种标记语言。
- (3) 文件结构: 许多文件都有其结构性,例如: E-mail 的发件人 (一个),收件人 (N 个),对方的电子邮件地址、电子书的章节,小节、文档的接收者,单位,日期等。传统的 HTML 对于文件的结构性并无强制性规范,而 SGML 则强烈要求文件的结构性,理由是这两种标记语言制定的目的不同,HTML 主要是为了文件的外观而制定的,而 SGML 不但为了文件结构而设计,并且本身就是一种描述语言,因此,SGML 利