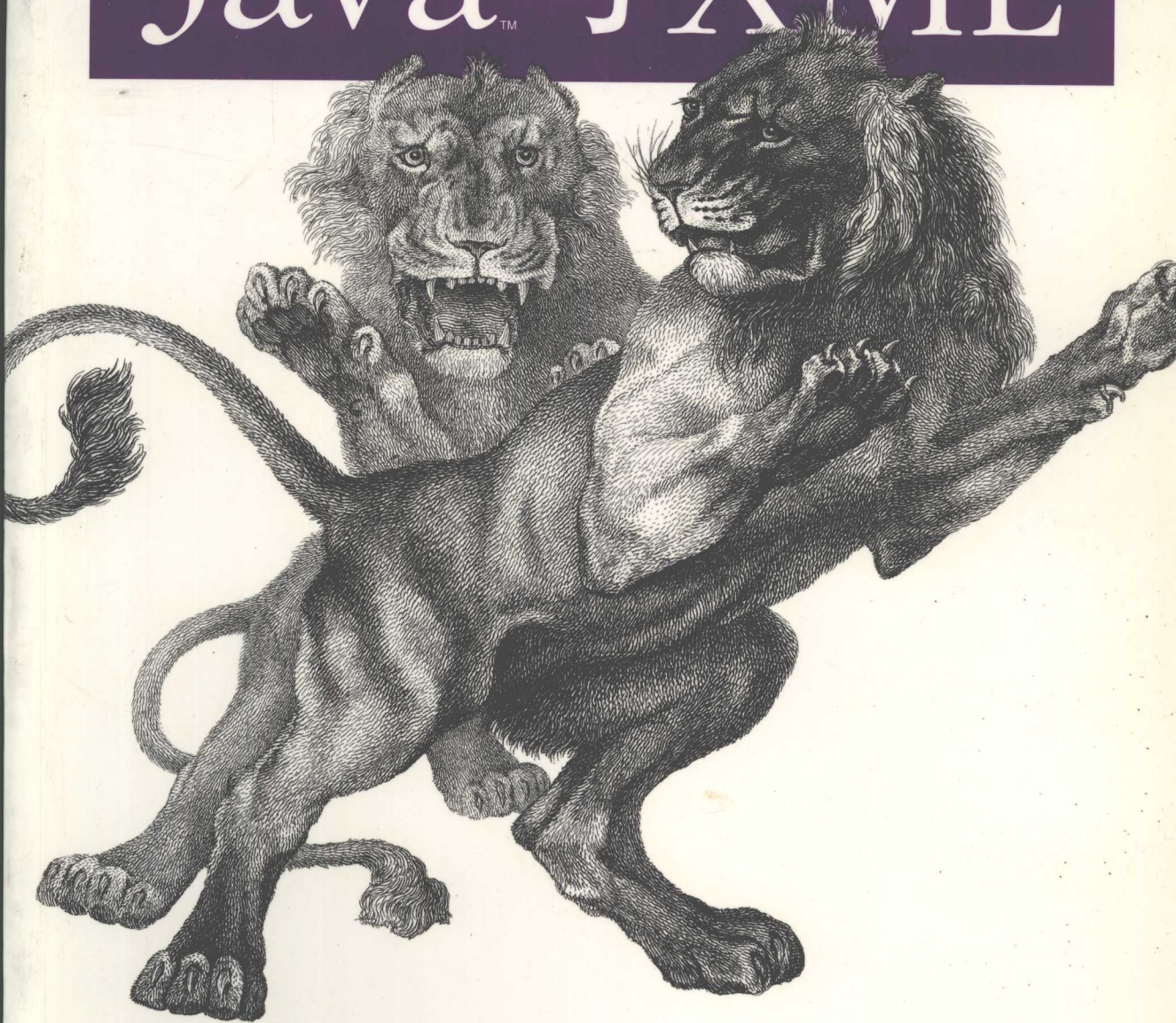


Java & XML

第二版

Java与XML



O'REILLY®
中国电力出版社

Brett McLaughlin 著

刘基诚 译

Java™ 与 XML

第二版

Brett McLaughlin 著

刘基诚 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Paris • Sebastopol • Taipei • Tokyo

O'Reilly & Associates, Inc. 授权中国电力出版社出版

中国电力出版社

图书在版编目 (CIP) 数据

Java™ 与 XML (第二版) / (美) 麦克劳林 (McLaughlin, B.) 著; 刘基诚译. - 北京: 中国电力出版社, 2003

(O'Reilly Java 系列)

书名原文: Java™ & XML, Second Edition

ISBN 7-5083-0946-4

I. J... II. ①麦... ②刘... III. ① JAVA 语言 - 程序设计 ② 可扩展语言, XML - 程序设计
IV. TP312

中国版本图书馆 CIP 数据核字 (2003) 第 027799 号

北京市版权局著作权合同登记

图字: 01-2003-0634 号

©2001 by O'Reilly & Associates, Inc.

Simplified Chinese Edition, jointly published by O'Reilly & Associates, Inc. and China Electric Power Press, 2003. Authorized translation of the English edition, 2001 O'Reilly & Associates, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

英文原版由 O'Reilly & Associates, Inc. 出版 2001。

简体中文版由中国电力出版社出版 2003。英文原版的翻译得到 O'Reilly & Associates, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly & Associates, Inc. 的许可。

版权所有, 未得书面许可, 本书的任何部分和全部不得以任何形式重制。

书 名 / Java™ 与 XML (第二版)

书 号 / ISBN 7-5083-0946-4

责任编辑 / 关敏

封面设计 / Ellie Volckhausen, 张健

出版发行 / 中国电力出版社 (www.infopower.com.cn)

地 址 / 北京三里河路 6 号 (邮政编码 100044)

经 销 / 全国新华书店

印 刷 / 北京市地矿印刷厂

开 本 / 787 毫米 × 1092 毫米 16 开本 31.5 印张 466 千字

版 次 / 2004 年 2 月第一版 2004 年 2 月第一次印刷

印 数 / 0001-4000 册

定 价 / 59.00 元 (册)

O'Reilly & Associates 公司介绍

为了满足读者对网络 and 软件技术知识的迫切需求,世界著名计算机图书出版机构 O'Reilly & Associates 公司授权中国电力出版社,翻译出版一批该公司久负盛名的英文经典技术专著。

O'Reilly & Associates 公司是世界上在 UNIX、X、Internet 和其他开放系统图书领域具有领导地位的出版公司,同时是联机出版的先锋。

从最畅销的《The Whole Internet User's Guide & Catalog》(被纽约公共图书馆评为二十世纪最重要的 50 本书之一)到 GNN (最早的 Internet 门户和商业网站),再到 WebSite (第一个桌面 PC 的 Web 服务器软件),O'Reilly & Associates 一直处于 Internet 发展的最前沿。

许多书店的反馈表明,O'Reilly & Associates 是最稳定的计算机图书出版商——每一本书都一版再版。与大多数计算机图书出版商相比,O'Reilly & Associates 公司具有深厚的计算机专业背景,这使得 O'Reilly & Associates 形成了一个非常不同于其他出版商的出版方针。O'Reilly & Associates 所有的编辑人员以前都是程序员,或者是顶尖级的技术专家。O'Reilly & Associates 还有许多固定的作者群体——他们本身是相关领域的技术专家、咨询专家,而现在编写著作,O'Reilly & Associates 依靠他们及时地推出图书。因为 O'Reilly & Associates 紧密地与计算机业界联系着,所以 O'Reilly & Associates 知道市场上真正需要什么图书。

作者简介

Brett McLaughlin 是 Java、XML、企业应用和开放源代码软件领域的资深权威。他是 Lutris Technologies 公司的 Enhydra 决策人员，负责 Enhydra 应用服务器的开发方向和决策制订。另外，他还是许多其他开放源代码项目的奠基人或合作奠基人，如 JDOM（目前在 Sun 的 JSR 中）、Apache Turbine（基于 Servlet 的 Web 应用框架）以及 Enhydra Zeus（XML 数据绑定框架）。另外他也是 OpenEJB、jBoss 以及 Apache Concoon 方面一言九鼎的人物，可算是 Java 和 XML “革命”的中流砥柱。

除了在技术上的贡献外，Brett 还是一个多产的作家；他是《Java & XML》（第一版）的作者，还是 IBM Java 与 XML 工具和技术新闻组的负责人，以及 *flashline.com* 半月刊的主要专栏作家，另外他还为 *IBM Developer Works*、*Java World* 和 *oreilly.com* 撰写了数十篇文章。

封面介绍

本书封面上的动物是狮子。这些“大猫”与其他单独行动的猫科动物不同，其区别在于它们要构成家族群体，称为狮群（pride）。狮群由 30~40 头狮子组成，其中大多数是雌性和她们的孩子。一头狮子的寿命大约为 3~4 年。成年的雄性狮子最多可以长达 10 英尺；惟一比它大的猫科动物只有老虎。狮子的眼睛对运动的事物尤其敏感，它从很远的距离就能够捕捉到其猎物的活动。在它的眼睛里有一些特殊的接收细胞，从而为之赋予了额外的夜视能力。

狮子生活在东非和南非，不过有些非洲狮的亚种已经濒临灭绝。亚洲狮（也称波斯狮）曾居住在印度、中东以及南亚。如今，其数量已经锐减至大约 290 只，在西印度 Gujarat 的 Gir 森林国家公园还可以看到这种狮子。地中海狮和南非狮都已绝迹。

狮子是食肉动物，并以大型群居动物作为捕猎对象。它们处在食物链的最上面，也是其所在环境的顶层猎食者。雌狮负责捕猎，而雄狮的角色是保护狮群不受其他敌

意雄狮的进犯。狮子跑动起来不如其他大型猫科动物（如印度豹）快。因此，其主要目标就是体态稍重、灵活性较差的动物，并且通过埋伏来捕获猎物，即将猎物驱赶至隐藏起来的狩猎成员处。它们采用协同、合作的技术，从而较之于单兵作战使得整个群体有更多的成功机会。非洲狮一般捕食角羚、斑马、羚羊、瞪羚、黑斑羚和长颈鹿。

目录

22	类野位容内
07	类野位器附
27	答籍教器
28	告野章不

29	只味器高 XAS 章四第
29	证
28	口器野位器具
29	器人器器器新世
29	口器野位器夏
104	答籍教器
108	告野章不

前言	1
----	---

第一章 简介	9
--------	---

XML 的重要性	9
XML 的重要组成部分	12
必备条件	14
下章预告	16

第二章 基本功	17
---------	----

基础知识	18
约束	28
转换	36
更多内容	44
下章预告	44

第三章 SAX	45
---------	----

做好准备	45
SAX 阅读器	48

内容处理类	54
错误处理类	70
疑难解答	75
下章预告	78
 第四章 SAX 高级知识	79
属性与特征	79
其他处理接口	86
过滤器和写入器	92
更多处理接口	99
疑难解答	104
下章预告	106
 第五章 DOM	107
DOM	107
串行化	112
可变性	125
疑难解答	125
下章预告	127
 第六章 DOM 高级知识	128
改变	128
名字空间	139
DOM Level 2 模块	142
DOM Level 3	156
疑难解答	159
下章预告	161
 第七章 JDOM	162
基础知识	162

PropsToXML	167
XMLProperties	178
JDOM 是标准吗?	189
疑难解答	190
下章预告	192
第八章 JDOM 高级知识	193
JDOM 核心剖析	193
JDOM 和工厂类	198
封装器	203
疑难解答	215
下章预告	218
第九章 JAXP	219
API 还是抽象	219
JAXP 1.0	221
JAXP 1.1	229
疑难解答	239
下章预告	240
第十章 Web 发布框架	241
框架的选择	243
安装	245
使用发布框架	249
XSP	266
Cocoon 2.0 及以后版本	281
下章预告	284

第十一章 XML-RPC	285
RPC 与 RMI	286
创建入门程序	289
给服务器加码	301
实战	317
下章预告	320
第十二章 SOAP	321
开始	321
配置	325
深入实践	329
继续深入	338
下章预告	346
第十三章 Web 服务	347
Web 服务	347
UDDI	349
WSDL	350
合练	353
下章预告	372
第十四章 内容协同发布	373
Foobar 公共图书馆	374
mytechbooks.com	384
推与拉	393
下章预告	402
第十五章 数据绑定	403
首要原则	404

Castor	410
Zeus	419
JAXB	427
下章预告	435
 第十六章 展望	436
XLink	436
XPointer	438
XML Schema 绑定	442
其他	443
下章预告	443
 附录一 API 参考	445
 附录二 SAX 2.0 特征和属性	485
 词汇表	489

前言

在一年多以前，写本书第一版前言的时候，对于所面临的大趋势，我的认识其实并不是太明确。那时我开玩笑地写到，XML会出现在人们的帽子和T恤衫上。现在，当我坐着写到这里时我身上正穿着一件写着大大的“XML”字样的T恤。不仅如此，我的帽子上也印着“XML”（事实上我有两顶这样的帽子！）。因此，XML的承诺已经实现了，这是毫无疑问的。一切真是太美妙了。

可是，这也意味着每天有更多的开发工作要做，而且 XML 正在以大大出乎我意料（做梦也想不到）的速度成长。这对 XML 是大好事。可本书的第一版就倒霉了：为什么一切似乎都过时了呢？我在是否要写 SAX 2.0 和 DOM Level 2 上还曾经犹豫过呢，它们现在却已成为业界标准；我在书中第一次引入了 JDOM，而现在它成为了 JSR（Sun 公司的 Java 标准请求过程）；我甚至没有提到 SOAP、UDDI、WSDL 和 XML 数据绑定，而新版本中它们却要占据整整三章！是啊，这世界变化真快。

如果你还在怀疑自己是否会在下几个月里用到XML,可以从本书得到帮助。如果你的电脑桌上正放着本书第一版,我推荐你看一看新版本。我想这对你仍然非常重要。这次我删去了对基本概念过多的描述,将XML的基本内容压缩成一章,并重写了几乎每个例子。我还加入了許多新的例子和章节。也就是说,我想让本书成为含金量更高的比较深入的技术书。这可能会使初学者花更多时间,因为帮助性的内容少了,但对于获取知识可能更有利。

本书的结构

本书的结构有些特别：前半部分（第一章~第九章）集中介绍有关XML和处理XML的Java核心API的基础知识。对三种操作XML的API（SAX、DOM和JDOM）中的每一个，都有一章讲基础，一章讲高级概念。第十章是个过渡章节，开始超出XML本身，讲述了JAXP、SAX和DOM之上的抽象层。本书的后半部分（第十一章~第十五章）集中阐释我所参与的各种会议和教程中频繁提到的各种具体的XML主题，让读者深入学习如何在程序中使用XML。这些主题包括关于SOAP和数据绑定的新章节，以及对B2B部分的更新。本书最后以两个附录结束。全书内容的摘要如下：

第一章 简介

介绍对XML各种宣传背后的真实技术，审视XML的种种术语，主要讨论为什么XML对于现在和未来的企业开发如此重要。

第二章 基本功

这是一堂介绍XML基础的速成课，从XML 1.0到DTD，从XML Schema到XSLT再到名字空间。对于已读过第一版的读者而言，本章相当于原来讲述XML的好几章的内容。

第三章 SAX

本章讲述SAX（Simple API for XML，用于XML的简单API），第一个介绍的处理XML的Java API。详述了解析全过程，以及SAX所捕获和开发者可使用的事件。

第四章 SAX 高级知识

本章进一步讲述SAX，讨论API中较少使用但功能强大的内容。可以从中找到如何使用XML过滤器将回调行为串联起来，如何用XML写入器和SAX输出XML，还能看到一些更少用的SAX处理方法，如LexicalHandler和DeclHandler。

第五章 DOM

本章沿XML知识体系继续前进，讲述另一个Java和XML API——DOM（Document Object Model，文档对象模型）。将学习DOM基础，了解当前标准（DOM Level 2），如何读写DOM树等等内容。

第六章 DOM 高级知识

本章进一步讲述 DOM，了解各种 DOM 模块，如 Traversal、Range、Events、CSS 和 HTML。我们还会了解新版本 DOM Level 3，讲述如何使用新功能。

第七章 JDOM

本章介绍 JDOM，讲述它与 DOM 和 SAX 的异同，以及如何用它来读写 XML。

第八章 JDOM 高级知识

本章更进一步审视 JDOM，讲述此 API 的实际应用，如何使用工厂类处理自己的 JDOM 子类，还有与 JAXP 的集成。还可以在实际中看到 XPath 如何与 JDOM 紧密配合。

第九章 JAXP

现在，作为功能完整的支持解析和转换的 API，JAXP 可以独立成章了。这里，我们讨论 1.0 和 1.1 版，完全学习如何使用 API。

第十章 Web 发布框架

本章讲述 Web 发布框架是什么，有何重要意义，如何选择。然后讨论 Apache Cocoon 框架，深入审视其功能，探讨它如何为 Web 上的动态内容提供服务。

第十一章 XML-RPC

本章讨论 RPC (Remote Procedure Call，远程过程调用)，它在分布式计算中的重要性 (与 RMI 比较)，XML 怎样使 RPC 成为某些问题的可行解决方案。然后讲述如何使用 XML-RPC Java 库，以及如何创建 XML-RPC 客户程序和服务器程序。

第十二章 SOAP

本章讨论如何使用 XML 格式的配置数据，为什么 XML 格式对跨平台应用如此重要，尤其是它与分布式系统和 Web 服务的关系。

第十三章 Web 服务

继续讨论 SOAP 和 Web 服务，本章详细讲述两项重要技术——UDDI 和 WSDL。

第十四章 内容协同发布

继续 B2B 应用程序的讨论，本章介绍如何使用内容协同发布 (content syndication) 实现企业程序之间的互相操作。将了解到 RSS (Rich Site Summary)，创建信息频道，甚至还会看到一些 Perl 代码。

第十五章 数据绑定

本章讨论一个高层 Java 与 XML API——XML 数据绑定。将了解数据绑定是什么，它是如何使处理 XML 变得不费吹灰之力的，以及目前已有的产品。我们将讨论三个框架：Castor、Zeus 和 Sun 公司的 JAXB (Java Architecture for XML Data Binding, Java XML 数据绑定架构) 的早期获取版本 (early access release)。

第十六章 展望

本章将对一些出现在“地平线上”的有趣新技术进行介绍，以开拓读者视野。有些技术可能会是昙花一现之物，而另一些将成为未来重要的角色。

附录一 API 参考

本附录详述 SAX、DOM、JAXP 和 JDOM API 中可用的所有类、接口和方法。

附录二 SAX 2.0 特征和属性

本附录详述 SAX 2.0 解析器实现中的特征和属性。

读者对象

本书以这样的信念为前提：XML 将很快成为 (某种程度上已经成为) Java 编程中不可或缺的部分。除第一章之外的其他章节将指导你如何使用 XML 和 Java，而不再集中讨论是否应该使用 XML。Java 开发者毫无疑问都要应该使用 XML。因此，如果你已经是 Java 程序员，想成为 Java 程序员，要管理 Java 程序员，或者工作与 Java 项目有关，本书都适合你。如果想更上一层楼，提高开发水平，编写更清晰的代码，或者想使项目成功地按进度和预算完成，如果你需要存取原有的老数据 (legacy data)，分布系统组件，或者只是想了解 XML 的众多宣传都在讲些什么，本书也适合你。

我尽量对读者不做过多假定，我不相信 XML 的门槛太高很难入门的说法。但是，我也相信，读者买本书是想学习基础之外的更多东西。因此，如果以前没有编写过 Java 程序，语言还需要入门，应该在本书之前读一下 Pat Niemeyer 和 Jonathan Knudsen 的《Learning Java》(O'Reilly) (译注 1)。读者无需了解 XML，但是我假定读者希

译注 1：中文版《Java 语言入门》已经由中国电力出版社出版。

望努力学习迅速掌握，因此我们会很快地过一下基础部分，大部分篇幅都留给高级概念。除非必要，我是不会重复素材的，因此在需要使用以前的概念时，可能得重读前面的章节，或来回跳跃地看。如果你已经会一些 Java，想学 XML，并准备在喜爱的编辑器中输入一些示例代码，从而成功掌握全书的内容，应该没有太大问题。

软件和版本

本书讲述了 XML 1.0 和到 2001 年 7 月为止的各种 XML 相关技术的最新版本。因为许多讲到的 XML 标准都不是最终版本，本书所讲述和读者读到本书时最新的标准可能会有一些差异（译注 2）。

书中的所有 Java 代码都是基于 Java 1.2 平台的。如果现在还没有开始使用 Java 1.2，我强烈建议你升级。单单是集合类就值得这样做。Apache Xerces 解析器，Apache Xalan 处理程序，Apache SOAP 库和 Apache FOP 库都是自 2000 年 6 月以来没有变化的最新稳定版本，而所用的 Apache Cocoon Web 发布框架是 1.8.2 版。所用的 XML-RPC Java 库是 1.0 beta 4 版。所有软件都是可以自由（免费）得到的，能从 <http://java.sun.com>、<http://xml.apache.org> 和 <http://www.xml-rpc.com> 下载。

书中例子的源代码基本上都是完整的。所有例子的源代码和二进制形式（包括丰富的 Javadoc）都可以从 <http://www.oreilly.com/catalog/javaxml2/> 和 <http://www.newInstance.com> 下载。所有实例都可以作为 Servlet、或者转换为 Servlet 运行，可以到 <http://www.newInstance.com> 在线查看和使用。

排版约定

本书使用以下英文字体。

斜体 (*Italic*) 用于：

- Unix 路径名、文件名和程序名。

译注 2：本书中文版付梓时，主要涉及的版本改动有：JAXP 1.2 已经发布，DOM Level 3 的核心标准已经发布。

- Internet 地址，比如域名和 URL。
- 黑体 (**Boldface**) 用于：
- GUI 元素的名字：窗口名、按钮和菜单选项等等。
- 等宽字体 (Constant Width) 用于：
- 应该原样输入的命令行和选项。
- Java 程序中的名字和关键字，包括方法名、变量名和类名。
- XML 元素名和标签、属性名，以及其他按文档中的形式显示的 XML 构造。

建议与评论

本书的内容都经过测试，尽管我们做了最大的努力，但错误和疏忽仍然是在所难免的。如果你发现有什么错误，或者是对将来的版本有什么建议，请通过下面的地址告诉我们：

美国：

O'Reilly & Associates, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

中国：

100080 北京市海淀区知春路 49 号希格玛公寓 B 座 809 室

奥莱理软件（北京）有限公司

询问技术问题或对本书进行评论，请发电子邮件到：

info@mail.oreilly.com.cn

另外，本书有一个网址，其中列出了实例、勘误以及未来版本的计划。可从以下页面访问：

<http://www.oreilly.com/catalog/javaxml2/>