

XML 精要:

语法详解 与编程指南

第二版

[美] Sandra E. Eddy & B. K. DeLong 著

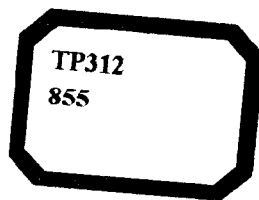
赵洪利 朱诗兵 等译



清华大学出版社
<http://www.tup.tsinghua.edu.cn>



XML 精要:



语法详解与编程指南

第二版

[美] Sandra E.Eddy & B.K.DeLong 著

赵洪利 朱诗兵 等 译

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书是一部全面、准确、权威讲解 XML (Extensible Markup Language) 语言及用法的最新专著。全书分为 3 大部分: 第 1 部分是 XML 参考, 第 2 部分是样式表参考, 第 3 部分是 XML 教程。

本书编写严谨, 结构清晰, 内容权威, 科学规范, 解释详尽, 语言通俗, 使用便利, 查找方便, 具有较高的参考和使用价值。对于 XML 协议研究人员、Internet 应用开发人员、计算机软件编程人员、高校相关专业师生, 是一部优秀的参考手册和指导著作。

XML 精要: 语法详解与编程指南 第二版

[美]Sandra E. Eddy & B.K. DeLong: XML in Plain English 2nd Edition

EISBN:0-7645-4744-5

Copyright©2001 by The Eddy Group, Inc.

Authorized translation from the English language edition published by Hungry Minds

All rights reserved. For sale in the People's Republic of China only.

北京市版权局著作权合同登记号 图字 01-2000-3532 号

本书中文简体字版由美国 Hungry Minds 出版集团授权清华大学出版社在中国境内出版发行。

未经出版者书面许可, 任何人不得以任何方式复制或抄袭本书的任何部分。

版权所有, 翻印必究。

本书封面贴有清华大学出版社激光防伪标签, 无标签者不得销售。

书 名: XML 精要: 语法详解与编程指南 第二版

译 者: 赵洪利 朱诗兵 等

出 版 者: 清华大学出版社 (北京清华大学学研大厦, 邮编 100084)

责任编辑: 林庆嘉

印 刷 者: 清华大学印刷厂

发 行 者: 新华书店总店北京发行所

开 本: 787×960 1/16 **印张:** 44 **字数:** 976 千字

版 次: 2002 年 4 月第 1 版 2002 年 4 月第 1 次印刷

书 号: ISBN 7-302-05183-6/TP·3043

印 数: 0001~3000

定 价: 68.00 元

关于作者

Sandra Eddy 从事计算机、Internet 和 Windows 应用方面的写作已经将近 20 年了。IDG Books 公司出版了她著作的 HTML In Plain English (第二版)、XML In Plain English 和 XHTML In Plain English。

B.K.DeLong 是 Web 标准咨询 ZOT Group 的首席研究人员, 计算机安全站点 Attrition.org 的工作人员, Web 标准工程指导委员会的总体 Web 标准质疑成员。他和妻子生活在马萨诸塞州北岸。他最近和 Simon St.Laurent 合作为 IDG Books 公司写了 Moving Toward XML 一书。

译者序

《XML 精要：语法详解与编程指南》是一部全面、准确、权威讲解 XML（Extensible Markup Language）语言及用法的最新专著。全书分为 3 大部分。第 1 部分是 XML 参考，主要包括 XML 简明参考、XML 列表、XML 语法、XLink 语言、XPointer 语言、XPath 语言等章节，重点讲解 XML、XLink、XPointer 或 XPath 的项、元素、属性或函数的名称、目的、完整语法和组件、使用说明以及相关项和属性，等等；第 2 部分是样式表参考，主要包括样式表简明参考、样式表列表、样式表分类、XSLT 组件、XSL 样式表语法等章节，重点讲解使用样式表将格式应用到 XML 文档，每一个 XSL 的性质、目的、完整语法、使用说明和相关性质，等等；第 3 部分是 XML 教程，主要包括 XML 概述、DTD 概述，创建 DTD，创建一个简单的 XML 文档，使用客户定制的 XML 元素，为 XML 文档添加链接和指针，使用 XSL 设计文档等章节，本部分旨在帮助和指导读者如何使用 XML 进行编程，重点讲解了 XML 文档结构和输出、内部和外部 DTD、根元素和它的后代元素以及属性、属性类型、属性值和各种类型的实体的定义、列表、表和窗体的基础知识、链接和扩展指针，利用可扩展样式表语言（XSL）进行转换，格式化和增强 XML 文档输出以及使用嵌套的 DSSSL Online（DSSSL-O）规则，等等。最后还附有 4 个附录，分别为 Unicode 字符和字符集、国家（地区）编码、语言编码和 EBNF 参考，提供的这些补充信息能帮助读者更好地理解 XML，更好地了解相关的支持技术背景。

本书编写严谨，结构清晰，内容权威，科学规范，解释详尽，语言通俗，使用便利，查找方便，具有较高的参考和使用价值。对于 XML 协议研究人员、Internet 应用开发人员、计算机软件编程人员、高校相关专业的师生，是一部优秀的参考手册和指导著作。

本书由赵洪利、朱诗兵、曹延华、蒋太杰、常冰、丁红勇、穆道生、李迎春等翻译，最后由赵洪利、王擎天、高理基统稿审定。翻译过程中得到了刘作学、吴刚、李长青、周江俊、赵军锁的大力支持和帮助，在此深表感谢。

译者

2001 年 10 月于北京

前言

欢迎使用本书。这本参考手册和教程为你提供了关于可扩展标记语言（**Extensible Markup Language: XML**，用来创建自定义的标记语言）、**XLink** 语言和 **XPointer** 语言（使用它们可以创建完善的 **XML** 文档）快速、全面的参考信息。另外，还有一部分和一章的内容讲述了可扩展样式表语言（**Extensible Stylesheet Language, XSL**），一个附录讲述了通用字符集（**UCS** 或 **Unicode 2.0**）及其所支持的字符集。书中还有很多示例，可以用来作为模板开发文档或作为学习的工具。

本书讲述了 **XML** 的核心内容。当你需要创建新的元素时，不需要去学习 **XML** 的背景知识；你所需要的只是项（**production**，即 **XML** 指令集或组成这些指令的组件）的名称、目的、语法、属性和实例。尽管本书没有包括 **Web** 编程和脚本编写，但它将引导你达到这样的水平。因此本书适合从初学者到高水平的所有开发人员。

本书是一本重要的 **XML** 手册，它应该被放在紧靠你的个人计算机的地方。

本书是如何组织的

本书易于使用，不管你的编程经验如何。

本书由 3 部分——第 1 部分“**XML 参考**”；第 2 部分“**样式表参考**”；第 3 部分“**XML 教程**”——和附录组成。

如果你想使用教程的综合参考和交叉引用，可以浏览第 1 部分：**XML 参考**，这部分内容包括下面几章：

- **XML 简明参考**。如果你知道要完成的任务，但是记不住 **XML**、**XLink**、**XPointer** 或 **XPath** 的项、元素、属性或函数的名称，可以查阅表中的相关任务并找到相应的在参考或教程部分中的详细内容。
- **XML 列表**。如果你知道要使用的项或属性的名称，则可以浏览这个具有简要描述、按字母顺序排列的列表。当找到某一个组件以后，则可以查阅在参考或教程部分中的相应内容。
- **XML 语法**。如果你知道项或属性的名称，但是还想学习更多的语法内容或某个组件的用法，可以参考本章。在这里，你可以找到每个项或属性、项的编号（经常出现在项的名称右边的上标中的数字，它是对 **XML** 规范的交叉引用）和它的目的、

完整语法和组件、使用说明以及相关项和属性。

- **XLink 语言。**XLinks（或扩展链接）与 XPointers 一起在 XML 文档中建立简单或扩展超链接。本章是对组成 XLink 的所有组件的全面参考。
- **XPointer 语言。**XPointers（或扩展指针）集中于 XLink 超链接的位置。本章是 XPointer 的项和函数及组件的全面参考。
- **XPath 语言。**XPath 语言支持 XPointer 和 XSL 转换(XSLT)。本章全面包括了 XPath 的项和函数及其组件。

在大部分的个人计算机应用中，创建字处理文档的个人已经使用样式表来设置标准段落格式。他们曾经在新、老文档的样式表和段落中应用各自的性质和样式。如果你想使用一个包括许多对教程的交叉引用的完整参考，请学习第 2 部分：样式表参考。这部分包括下面几章：

- **样式表简明参考。**如果你知道要如何设计一个元素，但是记不住 CSS、XSL 或 XSLT 的性质的名称，可以查看简明列表。
- **样式表列表。**如果你知道要使用的样式表的性质的名称，则可以浏览这个具有简要描述、按字母顺序排列的列表。当找到某一个性质以后，则可以查阅在参考或教程部分中的相应内容。
- **样式表分类。**如果你知道要使用的样式表组件的类别（性质、虚元素、虚拟类、at-规则、项、函数或格式化对象），则可以在这个按照类别组织的列表中找到它。找到该组件之后就可以查阅在参考或教程部分中的相应内容。
- **XSLT 组件。**你可以结合 XSL 样式表（见下面一段），使用 XSLT（转换）语言转换和设计你的 XML 文档。XSLT（转换）语言处理转换文档的任务，XSL 使用声明创建输出。本章是 XSLT 的全面参考。
- **XSL 样式表语法。**你可以使用样式表将格式应用到 XML 文档。在这一章中，你可以找到每一个 XSL 性质以及它的目的、完整语法、使用说明和相关性质。

如果你想学习 XML 和与之相关的一些技术，可以学习第 3 部分：XML 教程，其中含有对第 1 部分的交叉引用，这部分包括下面几章：

- **第 1 章 XML 概述。**这一章介绍了 XML 和其他的标记语言，讲述了超文本的历史，并在 XML 的近亲 SGML、HTML 和现在的 XHTML 的上下文中介绍了 XML。最后重点讲了 XML 文档结构和输出。
- **第 2 章 DTD 概述。**文档类型定义（document type definitions, DTDs）可以用来定义元素及其属性、实体和说明。DTD 使你可以逐渐地或有力地引导 XML 文档开发人员。这一章介绍了内部和外部 DTD。
- **第 3 章 创建 DTD。**本章告诉你如何创建 DTD——既可以从当前文档也可以从一个分离的文档进行。你将在本章学习如何声明根元素和它的后代元素，以及如何定义属性、属性类型、属性值和各种类型的实体。

- 第4章 创建一个简单的 XML 文档。当你学习了一些 XML 和 DTD 的知识以后，就可以创建自己的第一个 XML 文档了。你将在本章学习到如何将结构和内容添加到一个文档中。
- 第5章 使用客户定制的 XML 元素。一个简单的 XML 文档由文本、链接和偶然出现的图像（如果你的 XML 处理器支持图像）组成。要使文档更引人注目，你还要学习如何添加列表和表。为了更好地与你的网页的访问者进行交流，添加交互窗体是一个好的方法。本章包含了列表、表和窗体的基础知识。
- 第6章 为 XML 文档添加链接和指针。HTML 和 XML 文档的主要区别在于链接。HTML 文档支持一种类型的链接——简单链接。XML 提供了从简单类型到扩展类型的多种链接。本章概述了链接和扩展指针（它可以使你精确指向链接的目标）。
- 第7章 使用 XSL 设计文档。可扩展样式表语言（XSL）使文档书写和编辑人员可以用它转换、格式化和增强 XML 文档输出，使用嵌套的 DSSSL Online（DSSSL-O）规则。本章概述了基于当前 XSL 工作草案的 XSL 样式表，并给出了示例。

附录中提供的补充信息可以帮助你建立 XML 知识，更好地了解相关的支持技术的背景：

- 附录 A Unicode 字符和字符集。XML 注重国际化，在文档中使用多种语言。当你使用多种语言创建文档时，你所用到的不同的字母和符号都应该是 Unicode 所支持的。附录中提供了字符和字符集的说明表，特别是那些说英语的人常用的字符，并且指定了非英语的字符和字符集。
- 附录 B 国家（地区）编码。当你为国际用户创建文档或文档的一部分时，有时必须指定一个国家（地区）编码，它用来命名一种语言的某一国家（地区）的版本。该附录包含一个国家（地区）编码表。
- 附录 C 语言编码。当你为国际用户创建文档或文档的一部分时，必须为使用双字母语言编码的 XHTML 文档明确命名一种语言。附录中包含一个语言编码表。
- 附录 D EBNF 参考。XML 1.0 建议使用 Extended Backus-Naur Form（EBNF）说明来定义 XML 文档和 DTD 语法。要开发一种自定义语言或要理解声明元素和属性的 XHTML 文档类型定义（DTD），就需要懂得 EBNF。附录中含有语法和符号表。

本书中使用的语法

书中的每个项提供了 EBNF 说明和标准编程语法两种语法。

Extended Backus-Naur Form（EBNF）说明

XML 1.0 建议使用 Extended Backus-Naur Form（EBNF）说明来定义 XML 文档和 DTD

语法。开发人员可以书写一个 DTD 来为一个文档或文档集合指定元素、属性、实体和特殊字符。DTD 还设置每个组件的规则、限制和值。EBNF 使用某些约定：

- **Symbol|symbol ::=expression**: Each statement 是一个规则。在 XML 语法中定义一个项或符号。XML 是区分大小写的；开头具有大写字符的符号表示一个规则表达式（即一种组合字符或选项的方法），其他的符号都使用小写字符。

说明

当向文档中插入一个 XML 规则时，不要输入该符号或 ::=；应该只输入表达式。

- **#xN**: 输入 #x 和 N，一个符合任何 ISO/IEC10646 中的 UCS-4 编码值的十六进制整数。更多关于 Unicode 字符的信息，可参见第 1 章“XML 概述”、附录 A “Unicode 字符和字符集”和“Webliography”。
- **[a-zA-Z], [#xN-#xN]**: 输入 a~z、A~Z 或 #xN~#xN 范围中的一个字符。本书中的方括号 ([]) 较大，它用来标定那些包括在一部分表达式中的其他字符。
- **[^a-z], [^#xN-#xN]**: 不要输入任何与“NOT”字符接近的字符范围。
- **[^abc], [^#xN#xN#xN]**: 不要输入任何与“NOT”字符接近的字符。
- **"string"**: 输入括在双引号中的文字字符串。不要在一个表达式中将双引号 (") 和单引号 (') 混合使用。
- **'string'**: 输入括在单引号中的文字字符串。不要在一个表达式中将双引号 (") 和单引号 (') 混合使用。
- **(expression)**: 输入一个表达式，这个表达式由前面列出的部分 XML 语法和下面的语法（A 代表一个表达式）的组合构成。
 - **A?**: 后面跟着一个问号的表达式表示该表达式是可选的。
 - **AB**: 一个表达式后面跟着另一个，表示它们必定是完全匹配的。
 - **A|B**: 用管道符分隔的表达式表示“或者”。选择一个表达式或另一个。换句话说，只选择一个。在本书中的管道符较大，以便与项中的管道符区分开来。
 - **A-B**: 第一个表达式必须存在，减号后面的表达式必须不存在。

说明

一个范围符号（如 A-B）不含有空格，但是表示不存在的表达式时，(A - B) 的减号前后各有一个空格。

- **A+**: 该表达式必须出现一次或多次。
- **A***: 该表达式可能出现一次或多次。

说明

本书中包含一个范围的括号较大，它用来标定那些包括在一部分表达式中的其他字符。

标准编程语法

标准编程语法使用下列约定。

- {}：你必须选择一个包括在其中的属性、值、字符或标点符号。本书中的大括号较大，它用来标定那些包括在一部分表达式中的其他字符。一般说来，必需的属性列在可选的属性的前面。
- []：你可以选择一个或多个包括在其中的属性、值、字符或标点符号。本书中的括号较大以与在项中的括号区别开，它用来标定那些包括在一部分表达式中的其他字符。一般说来，必需的属性列在可选的属性的前面。
- |：管道符表示“或者”。选择一个属性或值，或者选另一个。换句话说，只选择一个。在本书中的管道符较大，以有别于项中的管道符。
- ^：不要选择任何这些字符，或与“NOT”字符接近的字符范围。
- ...：省略号表示无限延续前面的属性，后面的属性是这个序列的结束。
- *Italics*：斜体字文本代表你输入的一个变量（如文件夹/目录、文件名、路径、字符、数、URI 等）。大多数时候，使用双引号（" "）或单引号（' '）来包括一个变量，要保证以使用标准语法作为指导；绝对不要在同一个项中混用单、双引号。
- **default**：如果一个属性或值有下划线，表示它是默认的。换句话说，如果你不使用该属性，你的浏览器将自动使用默认的属性 and 值。

本书中使用的约定

在整本书中，每个项的描述都使用相同的通用格式。头部包括项的名称、它在括号中的上标编号（仅用于 XML、XLink 和 XPointer）和非常简要的描述。后面是更长的描述、两种类型的语法、关于组件和其他构成语法的属性的信息、该项的使用说明、一个或多个例子和一系列相关项。

说明

项的编号（如，^[84]）在参考 XML 规范或万维网联盟（W3C）工作草案时是很有用的。在 W3C 文档中，项是按照项编号与相关项或用来构成 XML 文档的组件组织的。

一些准许使用自己的输入项的属性和函数主要包含在本书第 1 部分的“XML 语法”和“XLink 语言”中。不含有项编号的属性和函数与含有编号的属性和函数使用相同的格式。本书中用到了下列约定：

- XML、XLink 和 XPointer 是区分大小写的，因此一定要正确输入大小写字符的组合。
- 要输入到 XML 文档中的项、属性和实体以等宽字体显示。
- 斜体字文本代表命名和输入的新的条目和变量（如文件名、数或 URL）。当看到斜体字文本时，用文件名、数或 URL 代替。

- 在 XML、XLink 和 XPointer 等章节中的例子，既有正常字体也有粗体字文本。粗体字用来突出显示例子中受当前项或元素影响的部分。
- 默认值用有下划线的文本表示，当你不使用某个属性时，它们将被自动提供。在你通读本书后，会发现下面两个图标几乎出现在每个章节中：

说明

说明图标用来引起你注意，其作用如同你对书上需突出的内容以标记符、笔或铅笔加以强调一样。

交叉使用

交叉引用图标指向本书中的其他章节，你可以在那里找到更多关于当前话题的信息。

如何与作用联系

我希望与你联系——特别是当你可以提供你曾经用来创建新的和值得称颂的 XML 文档的技巧、捷径和诀窍时。如果你发现我丢掉了 XML 的某个方面或重要的例子，一定要告诉我。我的电子邮件地址是 eddygrp@sover.net。

致 谢

不能回避这一点：在计算机或 Internet 书籍的封面上的作者的名字只是很多致力于这本书的问世并使之成为一个完整产品的人之一。没有编辑和专家们的支持，作者只能处于困境，完全迷茫，也只能重新杜撰一些陈词滥调。没有什么能像家庭和朋友的鼓励那样使人精神振奋——特别是在期限迫近的时候。阅读和试图理解技术标准是作者想在最后做的。我想在这里向所有曾经给我如此大力支持的人们表示感谢。

要向组稿编辑 Debra Williams Cauley 表示特别的感谢。／

还要向责任编辑 Andy Marinkovich 表示感谢，并向编辑 Michael Koch 表示衷心的、诚挚的谢意。

向其他使之变成有益的事情的 IDG Books 的人员表示感谢。感谢 S.B. Kleinman 所做的文字加工。

感谢技术编辑 B.K. DeLong 的 XML 专业知识和对细节的关注。

向我的代理人——Waterside Productions 的 Matt Wagner，表示感谢并致以最高的敬意。

感谢家庭和朋友们不断给我的鼓励。

感谢 Eli 和 Grace 的特别、不断的帮助，并永远记住 Indy 和 Toni。

最后要感谢本书的读者。希望大家能将你们对本书的看法和对下一版的改进意见告诉我。

目 录

关于作者	I
译者序	III
前言	V
致谢	XI
第 1 部分 XML 参考	1
XML 简明参考	3
XML 列表	10
XML 语法	13
XLink 语言	128
XPath 语言	146
XPath 语言	161
第 2 部分 样式表参考	211
样式表简明参考	213
样式表列表	234
样式表分类	242
XSLT 组件	252
XSL 样式表语法	301
第 3 部分 XML 教程	567
第 1 章 XML 概述	569
1.1 超文本的发展历史	570
1.2 标记语言的历史	571
1.2.1 SGML	571
1.2.2 HTML 和 XHTML	571
1.2.3 XML	572
1.2.4 XHTML	573
1.3 XHTML 和 XML	573

1.3.1	相似点	573
1.3.2	不同点	574
1.4	XML 文档结构	574
1.4.1	文档序	574
1.4.2	文档实例	576
1.5	XML 输出	579
第 2 章	DTD 概述	581
2.1	什么是 DTD	581
2.2	学习 XML 文档类型	582
2.2.1	格式良好的文档	582
2.2.2	有效的文档	583
2.3	理解内部和外部子集	583
2.3.1	内部 DTD 子集	584
2.3.2	外部子集	584
2.4	Extended Backus-Nuar Form 概述	586
2.5	理解一个 DTD	587
第 3 章	创建 DTD	589
3.1	声明一个 XML 文档	589
3.2	引用 DTD	590
3.2.1	内部 DTD	590
3.2.2	外部 DTD	590
3.3	定义根元素和它的子元素	591
3.4	指定元素出现的次数	592
3.5	声明子元素并列出它们的子元素	592
3.6	声明空元素 (Empty Element)	593
3.7	创建属性和值的列表	593
3.7.1	设置属性类型	593
3.7.2	设定属性为可选的或必需的	596
3.7.3	设定属性值	596
3.8	声明实体	597
3.8.1	解析的通用实体	597
3.8.2	字符和十六进制实体	598
3.8.3	参数实体	599
3.8.4	外部实体	599
3.8.5	数据实体	600

第 4 章 创建一个简单的 XML 文档	601
4.1 开始创建文档	601
4.1.1 声明 XML 的特性	601
4.1.2 指定文档类型、连接 DTD	602
4.2 建立文档结构	603
4.2.1 插入元素、起始标签和结束标签	603
4.2.2 使用空元素	603
4.2.3 添加属性和属性值	604
4.2.4 插入实体	605
4.3 HTML 文档转换为 XML 文档	607
4.3.1 考虑使用 DTDs	608
4.3.2 允许使用样式表	608
4.3.3 声明 XML 文档	608
4.3.4 命名根元素	609
4.3.5 遵守 XML 规则	609
4.3.6 转换属性和属性值	609
4.3.7 使用实体	610
第 5 章 使用客户定制的 XML 元素	611
5.1 使用列表	611
5.1.1 声明有序或无序列列表元素	611
5.1.2 声明定义-列表元素	613
5.1.3 列表嵌套	613
5.2 使用表格	613
5.2.1 声明表格元素	614
5.2.2 创建简单的表格	615
5.2.3 表格分段	615
5.2.4 行和列的组织	616
5.3 创建 XML 数据库	617
5.3.1 声明数据库元素	618
5.3.2 将数据库转换为 XML 文档	620
5.4 创建表单	620
5.4.1 学习控件	621
5.4.2 规划和设计表单	621
5.4.3 声明表单元素	622
5.4.4 定义表单	622

5.4.5 处理表单	623
第 6 章 为 XML 文档添加连接和指针	625
6.1 调用 XLink 的名称空间	625
6.2 关于 URIs	625
6.3 简单链接和扩展连接	626
6.3.1 简单链接	626
6.3.2 扩展连接	628
6.3.3 扩展指针	630
第 7 章 使用 XSL 设计文档	633
7.1 关于 DSSSL 和 DSSSL-O	633
7.2 XSL 简介	634
7.3 评估实例	636
附录	643
附录 A Unicode 字符和字符集	645
附录 B 国家（地区）编码	678
附录 C 语言编码	682
附录 D EBNF 参考	684

第 1 部分

XML 参 考

第 1 部分给出了 XML 及其相关技术(扩展链接(XLink)、扩展指针(XPointer)、和 XML 路径(XPath)语言)的综合参考。本部分共有 4 章主要内容——XML 语法、XLink 语言、XPointer 语言、XPath 语言,它们附有两个专门的表格内容。如果你知道要完成的任务,但是记不住项、元素、属性或函数的名称,可以查看“XML 简明参考”中的相关任务并找到相应的在第 1 部分和第 3 部分中的内容。另一方面,如果你知道元素或属性的名称,但想在参考或教程中找到它们,可以浏览“XML 列表”中按字母顺序列出的内容。

本部分包括

- XML 简明参考
- XML 列表
- XML 语法
- XLink 语言
- XPointer 语言
- XPath 语言