

XML: In Record Time

XML

轻松进阶

In Record Time

[美] Natanya Pitts 著
许菊芳 王雪松 译

轻松掌握，迅速提高

以实例为基础

以应用为目标



電子工業出版社

Publishing House of Electronics Industry

URL: <http://www.phei.com.cn>

XML: In Record Time

XML轻松进阶

〔美〕 Natanya Pitts 著

许菊芳 王雪松 译

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 提 要

XML是一项Web开发的新技术,利用它可以通过Internet或Intranet创建并发布信息。本书考虑到Web开发新老用户的需求,深入浅出地介绍了标记语言的基本概念、HTML语言的所有语法规则、XML的基本规则、XML组件、XML应用实例等。贯穿本书的程序实例清楚地表明了XML DTD和文档是如何编写的,提供了可直接使用的XML实例。本书还提供了有关Web开发的网址,集中起来也是一部关注XML发展的实时教材。

本书适合Web开发的初学者以及关注Web发展的有经验的老用户。



Copyright©1999 SYBEX Inc., 1151 Marina Village Parkway, Alameda, CA 94501. World rights reserved. No part of this publication may be stored in a retrieval system, transmitted, or reproduced in any way, including but not limited to photocopy, photograph, magnetic or other record, without the prior agreement and written permission of the publisher.

本书英文版由美国SYBEX公司出版,SYBEX公司已将中文版独家版权授予中国电子工业出版社及北京美迪亚电子信息有限公司。未经许可,不得以任何形式和手段复制或抄袭本书内容。

图书在版编目(CIP)数据

XML轻松进阶/(美)皮斯(Pitts. N.)著;许菊芳 王雪松译. —北京:电子工业出版社,2000.1

书名原文:XML: In Record Time

ISBN 7-5053-5550-3

I. X… II. ①皮… ②许… ③王… III. 计算机网络—可扩充标记语言XML—基本知识 IV. TP393

中国版本图书馆CIP数据核字(1999)第77240号

书 名: XML轻松进阶

著 作 者: [美] Natanya Pitts

译 者: 许菊芳 王雪松

责任编辑: 陈 宇

印 刷 者: 北京天竺颖华印刷厂

装 订 者: 三河金马印装有限公司

出版发行: 电子工业出版社 URL: <http://www.phei.com.cn>

北京市海淀区万寿路173信箱 邮编: 100036

北京市海淀区翠微东里甲2号 邮编: 100036

经 销: 各地新华书店

开 本: 787×1092 1/16 印张: 21.25 字数: 540千字

版 次: 2000年1月第1版 2000年1月第1次印刷

书 号: ISBN 7-5053-5550-3

TP·2809

定 价: 32.00元

版权贸易合同登记号 图字: 01-1999-1012

凡购买电子工业出版社的图书,如有缺页、倒页、脱页、所附磁(光)盘有问题者,请向购买书店调换。

若书店售缺,请与本社发行部联系调换。电话: 68279077

真心感谢DJ.为我所做的每一件事情。没有他，我不可能完成这本书。

致 谢

没有许多人的努力工作和支持，这本书就不会得以出版。首先要感谢我的父母**Charles**和**Swanya**，他们总是信任我；还要特别感谢我的亲密朋友**DJ**和**Jim**，是他们甚至在我不应该得到褒奖时，也在不停地鼓励我；感谢**LANWrights**公司的同事——**Ed Tittel**、**Mary Burmeister**、**Dawn Rader**，是他们在专业和个人方面的支持，帮助我度过了许多难关。

特别的感谢还要给**Sybex**出色的编辑们，是他们的艰苦工作，促成出版了这本高质量的书。他们是：**Krista Reid-McLaughlin**、**Diane Lowery**和**Kim Crowder**。同样也应该致谢**Sybex**生产部的**Nila Nichols**、**Shannon Murphy**、**Jeremy Crawford**和**Tony Jonick**。此外，还应该感谢**B. K. DeLong**，他是最优秀的技术编辑，任何作者都可以向他不停地询问任何问题。每个人的无私帮助，使得这成为了一本优秀的书。

最后需对**Robby**表示诚挚的谢意，**Robby**虽然只参与了这本书的最后过程，但是实际上他比我们任何人参与的时间都长。

译者序

由物理学家发明的Web技术（这确实令计算机专家汗颜），乘着Internet/Intranet的东风，正如火如荼地改变着我们的工作和生活方式。

要想使一种新技术获得广泛应用，得到大众的认可，宣传推广的作用不可磨灭。对于像XML这样的新兴技术而言，更是如此。

回顾Web技术的历史，从最早的HTML、DHTML到现在逐渐被人们重视的VRML、XML等，从时间来说并不长，但是其中蕴涵的技术飞跃却是巨大的。作为IT技术的传播者和见证人，我觉得自己有责任将国外最新、最好的技术介绍给国内广大的计算机爱好者。此外对于我个人，翻译的过程也是一个学习的过程。做一件事能够两全其美，何乐而不为呢？

由于XML是一项新兴技术，里面有许多专业词汇还没有统一的中文翻译标准，所以我只能参考其他相关的资料。如果其中有不准确的地方，请读者原谅。

在翻译的过程中，我力求尊重作者的原意，所以有些地方读起来可能并不十分顺畅，但是基本意思还是表达清楚了。

衷心感谢《计算机世界》报社的刘九如总编辑，是他一直在鼓励和鞭策我完成这次艰难的翻译之旅。

我还要感谢这些年来一直帮助我成长的兄弟姐妹们：永强、晓峰、海平、国良、易超、王雪、晶晶、李冬，还有可爱的小妹妹。

由于时间仓促，不当之处在所难免，敬请读者批评指正。

译者

1999年8月于北京

简介

XML是**Web**的未来。在以后的两年内，我们都将会对没有**XML**感到无所适从。媒体和技术官员们都在谈论**XML**的价值，将它说成是通过**Web**和**Intranet**交流或分发信息的革命，他们并没有夸大其词。

和其他**Web**新技术相比，**XML**在**Internet**开发人员的心里有着许多的疑惑：

- **XML**是一时的时尚？还是会顽强地生存下去？
- **XML**真会如描述的那样好吗（还是需添加一些东西才会更好）？
- **XML**能为我做些什么？
- **XML**的现实可行性如何？
- 想利用**XML**，需要哪些新的技能？需要哪些硬件和软件的配置？
- 现有的**XML**工具有哪些？哪些工具仍在开发当中？
- **XML**和**HTML**的关系如何？

通常我们都很难逃脱对任何新技术的过度宣传，以致于我们很难看到现实的真实情况。**XML**也是如此。

XML和上一代的**HTML**不一样，这种新的强大的标记语言，能够帮助处理数据，这些在以前都被认为是不可能的。**XML**本身并不能做到这些，要开发一个完整的**XML**方案，并不是仅仅需要一些自己创建的标志（**Tag**）和**Web**浏览器。

这本书的目的是揭开**XML**的神秘面纱，以便向读者展现**XML**的本质。当看完本书的最后一页，你就完全理解了什么是**XML**；它会如何影响**Web**的未来；需要做些什么才能完成**XML**方案；甚至还可以知道该如何用**XML**来创建自己的标记语言。

这本书为谁而写？

任何想利用**Web**技术、通过**Internet**或**Intranet**发布信息的人，都是本书的读者。**HTML**的熟练用户以及刚涉足该领域的新人，同样会发现包含在这本书每一页里的信息的价值。如果使用**HTML**已经有一段时间了，你会发现**XML**不仅仅是通过**Web**浏览器显示结构化文档的另一种方法，它还是描述内容的一个有力工具，这些内容可以被各种应用程序所处理。为了能正确而高效地使用**XML**，需要养成一些新习惯，甚至打破一些旧习惯。这本书就是帮助快速入门而写的，但是它不会强迫读者去学习那些已经知道了的信息。

书里讲了些什么

如果你是**Web**开发的新手，基本没有**HTML**经验，则本书最合你的胃口了。在进入未来的**Web**之前，本书的前几章将讲述**Web**的现状——**HTML**和**Web**页面开发。然后在**Web**页面设计和开发的基础之上，书中的其他章节将介绍**XML**的基本概念并帮助读者提高当前的工

作效率、为未来做好准备。

在本书的第一部分，你会遇到XML，了解当今Web标记语言——HTML和XML的基本情况。在第一部分里，将会介绍一些关键技能，提供深入使用HTML组件的基础；第二部分则讨论如何使用HTML组件；从第三部分开始，可以利用第一部分和第二部分学到的技能，考虑设计自己的XML文档；在第四部分，你有机会看到其他人在用XML做什么；最后在第五部分，将会介绍如何处理开发XML过程中遇到的问题。总之，书中所有的这些部分，都是为将XML作为全面信息分发方案而准备的。

本书的安排是这样的：先简要介绍HTML，然后给读者提供大量有关XML的知识，告诉读者一些XML的原理，此外还提供大量的练习。除了解释XML的基本概念之外，本书还用大量的篇幅探讨了如何实现XML。贯穿书中的程序实例，能让你清楚地看到XML DTD和文档是如何编写的，它们提供了可直接使用的XML的现实例子。当读完本书的时候，你就可以考虑把XML作为一个实际的方案，来解决各种各样的信息分发问题了。

本书还以下面的格式提供了附加信息：

说明：是用来指示需特别关注的地方的。比如，XML规范会经常发生变化，这样可以看看说明，关注一下在今后的几个月里规范可能会发生改变的部分。你还可以从说明里找到那些有用的Web站点。

提示：是为节省用户的时间而写的。提示是一些附加的帮助信息，也就是说，它们的目的是提供额外的信息。

警告：是为可能有的失误而准备的。

Web资源

作为给读者提供的便利，我们还将本书中所有列出来的URL集中到一起，建立了一个Web站点。这是一个特殊的XML软件网页，而且会经常更新有关信息。你可以在www.sybex.com里充分利用这个有用的在线资源。

XML的世界是变化不定的，这有点像玩过山车的游戏。尽管XML的未来不可预知，但是前进的过程是愉快而有趣的，我们期待着你能加入Web的未来之旅。

目 录

第一部分 XML介绍	1
第1章 了解XML	1
解答XML的常见问题	1
为XML做准备	7
小结	8
第2章 理解标记语言	9
标记语言的简单历史	9
HTML: SGML的一种	10
什么是标记语言?	11
将结构和显示区分开的好处	15
按规则办事的重要性	20
改掉不良习惯	21
小结	22
第3章 构成HTML的基础	23
Web: 一系列链接起来的文档	23
创建Web页面的基础知识	26
重要的HTML特性	31
从组件到组件	33
小结	33
第4章 用HTML标记描述文档	34
作为文档描述工具的HTML	34
HTML 4.0规范纲要: 分类的HTML标志	34
对HTML了解更多	61
从HTML到XML	62
小结	62
第5章 探究XML	63
XML主要部件介绍	63
初步了解XML规范的作用	69
继续研究	69
小结	69
第6章 理解基本的XML DTD和文档	70
DTD和文档如何协同工作?	70

基本DTD结构	72
基本文档结构	76
使用外部和内部DTD子集	82
设计完美的文档	85
小结	85
第7章 创建结构化的、有效的文档	86
结构化的和有效的文档之间的区别	86
该是输入标记的时候了	95
小结	95
第8章 使用不同类型的标记	96
为什么是两类标记?	96
实际例子	98
选择正确类型的标记	100
准备、设置、构建	101
小结	101
第二部分 XML的组件	103
第9章 为DTD设计元素	103
DTD中元素的作用	103
元素的不同类型	106
在DTD中声明元素	107
XML文档中的参考元素	109
构建更好的DTD: 元素计划	111
获取属性的细节	115
小结	115
第10章 给元素指定属性	116
DTD的属性作用	116
不同类型的属性	118
不同类型的属性值	120
在DTD中定义属性	122
为文档中的标志添加属性	125
构建更好的DTD: 属性计划	129
嵌套的规则: 内容模式	130
小结	130
第11章 为元素定义内容模式	131
嵌套的力量: 从HTML学到的教训	131
DTD中内容模式的作用	132

内容类型	134
在DTD中声明内容模式	135
在XML文档中坚持内容模式的准则	139
构建更好的DTD: 内容模式计划	141
实体: XML容器	142
小结	142
第12章 在DTD和文档中使用实体	143
DTD中实体的作用	143
在文档中声明实体	146
引用实体	154
通过链接将它们组合起来	159
小结	159
第13章 通过XLink和Xpointer链接	160
XML中Xlink和Xpointer的作用	160
在XML文档中创建XLink	163
当描述完成后: 配置XML文档	169
小结	169
第三部分 发布XML文档	171
第14章 处理XML文档	171
XML处理器的角色	171
解剖XML处理器	171
利用处理指令(PI)	174
可用XML处理器纵览	175
用CSS格式化XML	178
小结	179
第15章 用CSS添加样式	180
标记世界里样式单的角色	180
原始的样式单机制	181
在XML里使用CSS	181
构建CSS样式单	182
不同的属性族	185
爬上样式单的梯子	193
小结	193
第16章 使用XSL将XML转换成HTML	194
介绍XSL: XML样式单机制	194

	创建XSL样式单	195
	XSL样式规则示例	197
	写一个样式单, 将XML转换成HTML	198
	XSL的未来	199
	下一步: 当HTML是一个XML词汇表时	199
	小结	199
第17章	把现有的HTML文档变成有效的XML文档	200
	把HTML想像成XML词汇表	200
	把现成的HTML文档转换成有效的XML文档	204
	从理论到现实	215
	小结	215
第18章	用最新的Web客户浏览XML	216
	看看产业界对XML的态度	216
	理解在Web上浏览XML的真实性	217
	用Internet Explorer浏览XML	218
	在Netscape Navigator里浏览XML	221
	用HTML生成的Web站点: 这能够实现	223
	从一般到特殊: 研究个别的XML词汇表	224
	小结	225
第四部分	XML词汇表指南	227
第19章	为XML应用创建文档	227
	什么是XML应用?	227
	为内容选择正确的应用	228
	准备为现成的应用创建文档	233
	有效性检查	234
	修改XML应用	235
	一些XML DTD的深入探讨	235
	小结	236
第20章	用WIDL自动化Web站点	237
	什么是WIDL?	237
	WIDL主要结构的细节	240
	引导数据	242
	小结	243
第21章	用CDF引导数据	244
	什么是CDF?	244
	CDF主要结构的细节	250

用CDF通过Internet安装软件	260
小结	260
第22章 用OSD安装软件	261
什么是OSD?	261
OSD主要结构的细节	264
把OSD描述添加到CDF文件中	270
从安装软件到金融交易	271
小结	271
第23章 用OFX进行金融交易	272
什么是OFX?	272
OFX主要结构的细节	274
在多媒体中微笑	276
小结	277
第24章 用SMIL处理多媒体	278
什么是SMIL?	278
实现SMIL	281
SMIL主要结构的细节	283
超出词汇表的解决方案	287
小结	287
第五部分 现实的XML	289
第25章 判断XML是否是正确方案	289
三个关键问题	289
看看其他人怎样使用XML	291
在决定使用XML之后, 就去实现它	296
小结	296
第26章 处理实现XML过程中的问题	297
确定潜在的实施问题	297
把XML卖给老板	300
创建一个实施检验表	301
开发XML各方面的帮助工具	302
小结	302
第27章 选择适当的XML工具	303
XML工具箱	303
好的资源和好的工具一样重要	314
小结	314

第28章	发现最佳的XML资源	315
	记住： XML 是一项开发中的技术	315
	利用在线资源	316
	收集网络之外的XML资源	325
	结束语	326
	小结	327

第一部分 XML介绍

第1章 了解XML

- 回答XML的关键问题
- 理解XML作为一项Web技术的重要性
- 实现使用XML工作

可扩展标记语言（Extensible Markup Language, XML）架起了复杂的标准通用标记语言（Standard Generalized Markup Language, SGML）和有时显得功能有限的超文本标记语言（Hypertext Markup Language, HTML）之间的桥梁。在这一章里，我们将解答XML中的一些常见问题；探究XML到底是什么东西；XML作为一项Web技术的重要性；以及本书如何帮助你快速学习XML。

解答XML的常见问题

欢迎来到崭新而令人激动的XML世界，作为Web技术殿堂中一项崭新的技术，XML的出现引起了不小的轰动。当围绕着XML的热点快速升温，并且充满了许多诺言时，解答“什么是XML？”就显得不可缺少了。在阅读本书时，你可能已经知道使用XML能创建自己的标记，它将是下一个大型Web主导技术。这两种猜测都是属实的，但是对于XML，有比创建自己的标记更强大的能力。XML将铺设一条基于Web定制的文档描述和分发方案的道路，Web开发者再也不用将所设计的内容强迫转换成HTML格式了。一旦XML获得应用，它将改变我们设计和应用Web文档的方法。

尽管所有这些都是美好的，但是还有许多有关XML的问题难于琢磨或者难于找到答案。没有这些答案，你就不会知道XML是否是适合你的技术。我已经发现即使是最好的计算机图书，它们很好地解释了许多问题，提供了指南、示例代码和其他用户可能在开始这项技术时所需的帮助，但它们并不能真正地解释用户可能会遇到的基础问题。有着广泛多样性的FAQ——经常被问到的问题和它们的答案的列表——在许多基于Internet和Web的项目中可以看到，显示了某项技术的基本问题，这些问题确实需要在学习怎样利用它之前被回答。对于XML，也是如此。

XML的在线FAQ

XML的FAQ在Internet里才刚刚开始出现，它们中的大部分都将重点放在XML的某个特定的应用上。两个较好的XML FAQ是：

- 由Peter Flynn维护的W3C XML工作组的XML FAQ，网址是<http://www.ucc.ie/xml/>
- 由Microsoft维护的Microsoft XML FAQ，网址是<http://www.microsoft.com/xml/xmlfaq.htm>

Peter Flynn的XML FAQ有点偏重于技术，但是它定期更新。当你对XML的某些方面有疑问时，它确实是一个非常好的参考。这个FAQ由Flynn维护，属于World Wide Web Consortium (W3C) 的XML工作组，这个工作组是XML领域的权威，总是发布一些最新的官方FAQ；Microsoft XML FAQ比起Flynn的FAQ来，缺少一些广泛性，但是它容易阅读，并且包含了许多有用的信息。应该承认，这个FAQ是从Microsoft的观点来写的，里面包含了一些纯粹基于Microsoft的内容，它是Microsoft发布最新XML信息的固定来源。

这本书中除了介绍XML之外，我还收集了自己的XML FAQ，回答那些XML中最普通而最重要的问题。我的FAQ并不详尽，但它确实关注的是XML的关键问题，目的是驱散XML初学者心中的“疑惑”，为用户搭建一个坚实的XML基础，告诉他们怎样了解XML，怎样用XML来工作。此外，因为FAQ是为了突出主题中最重要的方面而设计的，在本书中的部分章节——从第6章“理解基本的XML DTD和文档”开始——都包括一个FAQ，阐述了本章中所提到的重要的主题和概念，这就是Q&A。

W3C：最终的权威

Internet和WWW构建了世界范围内普遍接受的标准，这些标准使得各种运行不同操作系统和软件的各式计算机可以互相通话。W3C是为了保证Internet和Web的标准不被某一个厂商或操作系统所独占而建立的。此外，W3C还负责保证现在的标准能不断被更新，以迎合Internet和Web发展的需要。W3C还负责组织业界的专家来发展新的标准。

有关HTML、图形、样式表 (Style Sheet)、Web服务器和其他许多Web技术的官方标准，已经是由W3C制定并维护着。W3C只可以推荐新的标准，并且希望厂商们能够生产出支持这些标准的产品。通常，如果某个公司希望它的产品销售得好，拥有大量的用户，则这个公司一定会生产那些支持由W3C所推荐的标准的产品。所有主要的Internet和Web软件和硬件开发者都是W3C的成员，他们自己直接制定标准，同时他们的产品也必须遵循这些标准。

W3C的站点 (<http://www.w3.org/>) 里包含了所有已经制定好了的标准文档，以及那些还在不断完善发展阶段的标准和初始信息。如果你工作时需要和Internet/Web技术打交道，但是还没有访问过W3C的站点，建议你去看一看，这个站点上的资源对任何人都是公开的。

什么是XML？

位于<http://www.w3.org/XML/Activity.html>的W3C XML Activity页面提供了这个问题的答案：

XML——可扩展标记语言——是一种基于SGML的、简单灵活的语言。尽管最早是为了解决大规模出版所面临的挑战，XML现在却被用来在Web上标记各种数据，起着越来越重要的作用。在XML的帮助下，人们不仅可以获得所需要的信息，而且可以得到Web上的XML元数据 (metadata) ——有关信息的信息——这将对许多基于Web的应用有帮助。

XML将使信息消费者和信息生产者之间的沟通变得更加容易，信息搜索和信息交换的许多任务能够在XML下实现自动化，XML将提供一个通用的框架（framework）来描绘信息，这将使每个人都能从中受益。

通俗地讲，XML就是一种创建标记语言的语言——或者是一种元标记语言，它尤其适合同种类型的内容。标记语言使用直接插入在文本中的标记来描述文本的各种块和部分。在下面的例子中，<PARA>和<ITALICS>标记描述了文章的一段和一个斜体单词：

```
<PARA>"XML - the eXtensible Markup Language - is a simple and very  
flexible language based on SGML. Although originally envisaged to meet  
the challenges involved in large-scale publishing, XML is set to play  
an increasingly important role in the markup of a wide variety of data  
on the Web. Not only will XML help people find the information they  
want, but the wealth of XML metadata on the Web - information <ITALICS>  
about</ITALICS> information - will help many Web-based  
applications.</PARA>
```

标记语言并不关注它所描述的内容是怎样的格式，相反它将注意力放在如何准确地描述它的内容上。HTML是一种标记语言，标记语言的概念对于理解和执行XML是一个关键。在本书中，从第2章“理解标记语言”开始，你将学到许多有关标记语言的知识。当完成第二部分后，你应该能够用XML来编写自己的标记语言了。

总之，XML是一种能够运行在任何平台和操作系统上的标记语言，是用来给开发者提供一种技巧，以使他们更好地描述内容。XML最早的目的是为了出版，但是现在已经发展到被用来使Web上交换数据变得更加容易和更加高效了。XML通过允许开发者们编写自己的DTD（文档类型定义）来实现。DTD描述了一系列标记和特性，这些标记和特性能够被用来描述特定种类的文本。DTD是一种标记语言规则，这种规则规定了哪种标记元素能够被用来描述一个文档。如果你想创建自己的标记，则首先必须在DTD中定义它们。第2章介绍了DTD；第6章在更深的层次描述了DTD。XML通过DTD定义的标记语言被称为XML字典（XML vocabulary）或者应用。

XML是在W3C工作组的支持下发展起来的，以保障XML的规范是标准的，不被某个厂商所独占。每个XML字典都有一个有特殊目的的某种类型的文本，用来直接描述XML规则的内容，所有的XML字典组成了XML字典子集。有两个XML字典的例子，它们是通用标记语言（Genealogical Markup Language, GedML）和化学标记语言（Chemical Markup Language, CML）。这两种标记语言都属于XML字典，而且已经根据XML规范得到了发展，但是它们有着不同的目的。GedML用来描述传统的数据，而CML却用来描述化学公式和分子式。这两种标记语言都是通过DTD来定义的，分别指定了可以用来描述传统信息和化学信息的元素。第6章包含了对DTD和特定的XML字典怎样协同工作的深层次探讨。通常XML提供了一系列标准化的规则，用来描述DTD及其用于Internet和Web的文档。你可以利用XML来编写定义自己的字典和文档的DTD，或者根据别人已经编写好了的字典中的DTD来创建文档。

XML除了能够描述内容之外，它还非常胜任用来描述元数据。元数据是用来描述信息的信息。一个我们都打过交道的元数据的例子是图书馆的目录卡，每一张目录卡（或者是计算机中的电子目录），它本身就是一个信息资源，提供了另外一个信息资源的信息，另外一

个信息资源通常是一本书、杂志或电影片段。连贯的描述元数据（如XML）的方法，会产生更加有组织性的Web资源目录，能够使得“Web居民”高效率地搜索成千上万的Web页面。

简单地讲，XML提供了更广泛的方法来描述文档内容，同时通过一种能够在所有计算机上运行、与平台和操作系统无关的方法，提供描述元数据的机制。XML利用适合特定种类内容的DTD所定义的字典，使得对Web数据的利用提高到一个新水平。

为什么XML会发展？

XML之所以能够获得发展，是因为文档设计者和内容专家都意识到，用HTML来处理所管理的许多任务没有太多的回旋余地。并不是所有的内容都可以被描述成段落、清单或表格。因为HTML唯一可用来描述Web内容，所以必须把所有的Web内容都压缩成段落、清单或表格等。如果你试着将一个正方形的钉子变成一个圆形的洞，就会知道为什么HTML的表现能力是如此的有限了。

W3C的各个工作小组的成员们都意识到，HTML不能容纳所有类型的数据，否则HTML会变得更笨重。因此不得不发展一系列的标记语言，每一种标记语言都有特殊的目的，但是他们没有任何办法预先知道每个人对标记语言的需求。即使回到SGML的时代也不能解决问题，因为SGML是一种具有“过度杀伤力”的解决方案。

XML为开发者提供了一种结构化的环境，它允许开发者把那些不适合HTML模型的内容放入DTD中。通过定义一种用来创建基于Web的标记语言的元语言（metalanguage），而不是整个的一个新的标记语言组的办法，W3C提供了一种为现存的Web结构进行定制的解决方案。

如果你对Web还很陌生，这些有关HTML和SGML的讨论使得你一头雾水，也先不要着急。第2章会介绍标记语言的基础知识，此外还有如何利用XML和HTML开展工作的基本方法。第3章“构成HTML的基础”和第4章“用HTML标记描述文档”则提供了HTML的应急课程，将帮助你创建第一个基于HTML的Web页面，并为下一步用XML定义Web分发的内容做准备。

XML和HTML有什么关系？

在许多方面，XML是HTML的“表妹”，甚至于将来可能会成为它的“父母”。XML和HTML都是由SGML派生而来的，但是XML是一种元标记语言，而HTML是一种特殊化的标记语言。SGML是所有标记语言的祖先，包括XML和HTML。在第2章中，你将看到这三种标记语言为什么会得到发展以及它们是怎样发展的，还有它们在信息描述和分发方面所体现的不同作用。实际上HTML有可能会成为一个XML字典，用来描述简单的文档并使Web站点向下兼容低版本的浏览器。

当开发者需要将一个非专用的标记语言和通过Internet交换的数据相结合时，HTML为XML铺设了一条道路。HTML在描述各种文档方面的广泛应用，导致了开发者在寻找解决方案时的需求。HTML是为了满足设在瑞士的粒子物理欧洲实验室的特殊要求而创建的，很快就发展成为了各种各样需求的解决方案。XML则是为了满足HTML广泛应用的需要而构建的。

如果你了解HTML的基本知识，则你也肯定知道一些XML。现在的Web暂时还是由HTML构建的，所以需要了解一些HTML才能执行你的XML方案。许多在XML字典中用来描述和