国外计算机科学经典教材

DEFINITIVE XML Schema

XML 模式权威教程

Priscilla Walmsley 著
陈维军 乔安平 英宇 译

- 透视XML 模式的强大功能
- 完全涵盖了已获广泛认可的、W3C推荐的XML 标准
- 全面、实际地设计 XML 模式
- 帮助富有经验的 DTD 开发人员迁移到 XML 模式
- 由 W3C XML 模式工作组的成员——Priscilla Walmsley 编写，具有极高的权威性

清华大学出版社
http://www.tup.tsinghua.edu.cn
Pearson Education

XML 模式权威教程

Priscilla Walmsley 著

陈维军 乔安平 英宇 译

清华大学出版社

北京市版权局著作权合同登记号：01-2002-3006

内 容 简 介

本书介绍的是一种功能强大的文档模式语言：XML Schema，它的独特功能包括强类型、模块化、继承以及一致性约束等。本书主要内容包括：XML Schema 如何为 XML 文档结构、内容和数据类型建模提供严格而完整的标准，XML Schema 的元素、属性和类型等构件，以及类型派生、模型组、替换组等高级内容。

本书适合于 XML 开发人员、系统架构师等专业人员及相关的计算机知识爱好者阅读。

图书在版编目(CIP)数据

XML 模式权威教程/(美)沃姆斯利著；陈维军，乔安平，英宇译.

—北京：清华大学出版社，2002

书名原文：Definitive XML Schema

ISBN 7-302-06096-7

Ⅰ.X… Ⅱ.①沃…②陈…③乔…④英… Ⅲ.可扩充语言，XML—程序设计 Ⅳ.TP312

中国版本图书馆 CIP 数据核字(2002)第 091354 号

出 版 者：清华大学出版社(北京清华大学学研大厦，邮编 100084)
http://www.tup.tsinghua.edu.cn

责任编辑：陈宗斌

封面设计：康博

版式设计：康博

印 刷 者：北京通州区大中印刷厂

发 行 者：新华书店总店北京发行所

开　　本：787×1092 1/16　**印张**：19.5　**字数**：499 千字

版　　次：2003 年 1 月第 1 版　2003 年 1 月第 1 次印刷

书　　号：ISBN 7-302-06096-7/TP・3541

印　　数：0001～3000

定　　价：46.00 元

目　　录

第1章　XML Schema简介

本章简要地介绍了模式(Schema)及其重要意义。同时也讨论了基本的模式设计目标，并介绍了现有的多种模式语言。

1.1　XML Schema 的概念

模式(schema)这个词语表示图解、计划或框架。在 XML 中，它指描述 XML 文档的文档。假使您有一个例 1-1 所示的 XML 实例，它包含了一个 product 元素和一个属性(effDate)，前者又有两个子元素(number 和 size)。

例 1-2 给出了一个描述这个实例的模式。该模式中包含了元素和属性声明，用于将数据类型和元素类型名赋予元素和属性。

例子 1-1　Product 实例

```
<product effDate="2001-04-02">
  <number>557</number>
  <size>10</size>
</product>
```

例子 1-2　Product 模式

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="product" type="ProductType"/>
  <xsd:complexType name="ProductType">
    <xsd:sequence>
      <xsd:element name="number" type="xsd:integer"/>
      <xsd:element name="size" type="SizeType"/>
    </xsd:sequence>
    <xsd:attribute name="effDate" type="xsd:date"/>
  </xsd:complexType>
  <xsd:simpleType name="SizeType">
    <xsd:restriction base="xsd:integer">
      <xsd:minInclusive value="2"/>
      <xsd:maxInclusive value="18"/>
    </xsd:restriction>
  </xsd:simpleType>
</xsd:schema>
```

1.2 模式的用途

1.2.1 数据确认

模式最常见的用法之一就是依照预定义好的规则来验证 XML 文档的有效性。模式可以确认：

- 元素和属性的结构。例如，一个 product 对象必须有 number 和 size，以及可选的 effDate(有效日期)。
- 元素的顺序。例如 number 必须出现在 size 之前。
- 元素和属性的数据值，根据范围、枚举以及样式匹配。例如，size 必须为一个 2 到 18 间的整数，而 effDate 必须是有效的日期。
- 实例中值的惟一性。例如，实例中所有的产品编号必须是惟一的。

1.2.2 交易双方的合约

通常情况下，XML 实例会在组织之间传递。模式就如同交易双方都必须遵循的合约。它清楚地说明了文档结构的规则以及要求。如果一个实例被模式验证过，那么“合约”就可以使用可用的工具来实施。

1.2.3 系统文档

模式可以为 XML 实例中的数据提供文档。任何需要了解这些数据的人，都可以从模式中获得关于项目的名称、结构和数据类型等信息。要包括附加文档，您也可以为模式中的组件添加注解。

1.2.4 数据扩充

模式处理过程也可以添加到实例中。它为元素和属性插入默认和固定的值，并根据数据类型规范化空白空间。

1.2.5 应用程序信息

应用程序在处理文档中的一些特殊类型时，模式提供了一种方式来为应用程序提供数据的附加信息。例如，您可以包含一些关于如何将 product 元素实例映射到数据库表的信息，并且允许应用程序使用这些信息来自动更新带有该数据的特定的表。

除了处理时可用之外，模式中的该信息也可以用于生成代码，比如：

- 用来修改该信息的用户界面。例如，如果已经知道 size 是介于 2 到 18 之间的数值，那么可以生成一个带有滑动条的界面，以这两个数作为滑动条的限制值。
- 将实例数据转换为如 XHTML 这样容易阅读的表现形式的样式表。例如，如果您知道 number 元素的内容有一个易于理解的名称为“Product Number”，那么就可以将它作为列标题。

- 从数据库中插入或提取数据的代码。例如，如果您知道产品编号对应于 PRODUCTS 表中的 PROD_NUM 列，那么您能够生成一个有效的例程来将它插入到该列。

一些工具已经开始利用模式的这种能力。在未来数年中，人们将以许多有创造性的新方法使用模式。

1.3　模式设计

XML Schema 功能多样，经常可以通过不同的方法来精确地描述同一事物。在设计模式时选择不同的方法可能影响到模式的可用性、准确性和适用性。因此，最重要的就是在创建模式时，要明确地知道自己的设计目标。这些目标可能会因为使用 XML 的方法不同而不同，但有些目标对于各种应用案例来说是通用的。

1.3.1　准确性和精确性

显而易见，模式应该能准确地描述 XML 实例，并且能够确保该实例的有效性。模式在描述数据时也应该是精确的。精确性能带来更完整的有效性，以及结构更好的文档。精确性可以通过定义真实代表有效值的受限数据类型来实现。

1.3.2　明晰性

模式应该非常清晰，允许读者直观地理解被描述实例的结构和特征。模式的明晰性可以通过以下方式实现：

- 选择合适的名称
- 命名规则的一致性
- 结构的一致性
- 良好的文档
- 避免不必要的复杂性

1.3.3　广泛适用性

仅仅是因为一些特殊的应用目的而创建模式吗？有时候，这可能是适用的。但最好使创建的模式有更广泛的适用性。例如，一家商业公司如果只处理国内账目，那么可能就不会将 country 元素声明作为 address 的一部分。出于一致性和将来应用的目的，他们应该考虑将它作为一个可选的元素添加进去。

模式的广泛适用性包含了两部分：可重用性和可扩展性。可重用的模式组件是模块化的和格式良好的，鼓励模式编写者在其他模式中重用它们。可扩展的组件则是灵活的和开放的，允许其他模式编写者在原有基础上根据需求进行扩展以备将来之用。由于可重用性和可扩展性非常重要，本书第 21 章的全部内容都是对它们的详细讲述。