

the method of XML document modeling by UML modeling tools

用 UML 建模工具为 XML 文档建模的方法



摘要: XML 作为一种具有严格结构的 WEB 符号标记语言, 越来越在电子商务中得到了更大应用。本文尝试了一种用 UML 建模工具为 XML 文档建模的方法, 用以提高 XML 的编写效率, 并给出了具体的实例。

关键词: XML UML DTD Schema 建模

1 为 XML 文档建模的必要性

XML 自从 1998 年 2 月正式诞生后, 就在世界上广受推宠, 不仅因为它具有 SGML 的可扩展性、结构化、有效性三种优势和 HTML 的简单性, 而且因为它具有 HTML 所不具有的严格结构化。它的这种结构化使其能够有效地通过 WEB 在应用程序间传递结构化的数据, 极大地拓展了电子商务的发展。

一个合理高效地 XML 文档一定要有一个合理的结构。如何来构造它的结构呢? 借助于当今日臻完善的 UML 建模技术可以设计出一个合理的结构, 而且运用建模技术中的逆向工程技术可以反复推敲 XML 文档的结构, 使其更加完善, 最终设计出合理、高效的 XML 文档。

2 XML 文档的结构

XML 使 WEB 文件内容 (数据) 和描述 (数据的表示形式) 分割开来。XML 的结构化, 简单地说, 就是 WEB 文件的数据的格式化, 也就是数据模型。一个 XML 文档相对它的数据模型就像一个对象相对于它的类。有两种方式来表示 XML 文档的数据模型。一是 DTD (文档类型定义: Document Type Definition) 文件, 二是 Schema 文件。

2.1 DTD

DTD 依靠特定的语法来描述 XML 文档的结构, 它用单独的字符如问号、星号以及加号来描述文件的内容模型。DTD 的基本部件是元素和属性, 它们确

定了XML文件的逻辑结构。我们可以把元素看作一个信息对象，而属性表示这个对象的性质。使用元素和属性前都要进行声明。

2.1.1 元素的声明格式如下

<ELEMENT ElementName Type>其中 ELEMENT 是XML的保留字。ElementName 是元素的名称。Type是元素的类型。XML只支持下列四种元素类型：

Empty 指元素内容为空，但它可以有属性

Element Only 指元素只可包含子元素

Mixed 指元素可包含子元素和字符数据的组合

Any 指元素可包含任何DTD允许的内容

在 Element only 元素类型中，用问号 (?) 来确定子元素已出现一次或一次都不出现。由星号 (*) 来确定子元素可出现任意次。用加号 (+) 来确定子元素至少出现一次。同时用圆括号封装一个子元素的序列或一个选择组。用管道符 (|) 来分开选择组中的各个选项。

在 Mixed 元素类型中，只包含字符数据的声明为 <ELEMENT Text (#PCDATA)> 其中 (#PCDATA) 用来表示是字符数据。

2.1.2 属性的声明格式如下

<!ATTLIST ElementName AttrName AttrType Default>

其中 ATTLIST 是XML的保留字。ElementName 是元素名称。AttrName 是属性名称。AttrType 是属性类型。Default 是属性类型的缺省值。AttrType 有字符串类、枚举类、标记类三大类。

2.2 Schema

Schema和DTD一样，也是为了描述XML文档的结构。比DTD高明地是：

(1) Schema使用了XML的语法，不用再学什么特定语法；

(2) 它将元素和数据类型联系了起来，而且元素的内容不仅仅限于字符串类型，也可以用非常特别的类型，比如整型、浮点型、布尔型、日期型等。

虽然Schema比DTD要更优，并且当前一些公司(如微软)已明确支持Schema，但它还只是一份草案，所以本文不加多述。

3 具体建模的过程分析

XML文档建模的核心就是如何在UML工具中合理、准确地表现DTD或Schema文件中要表示的数据结构，并最终生成DTD或Schema文件，或者由DTD或Schema文件逆映射到UML中去。

下面我们以建模工具Power Designer 8.0和网上图书订购系统中的一个订单为例。如图1所示，根据需求生成的类DingDan和DingDanMingXi的OOM (Object-Oriented Model) 图，当前的操作语言为XML-DTD

执行Generate XML...操作，生成的DTD文件如下：

```
<!ELEMENT DingDanHao (#PCDATA)>
<!ELEMENT DingShuRen (#PCDATA)>
<!ELEMENT SouLiYuanGong (#PCDATA)>
<!ELEMENT JinE (#PCDATA)>
<!ELEMENT DateTime (#PCDATA)>
<!ELEMENT DingDan (DingDanHao? | DingShuRen?
```

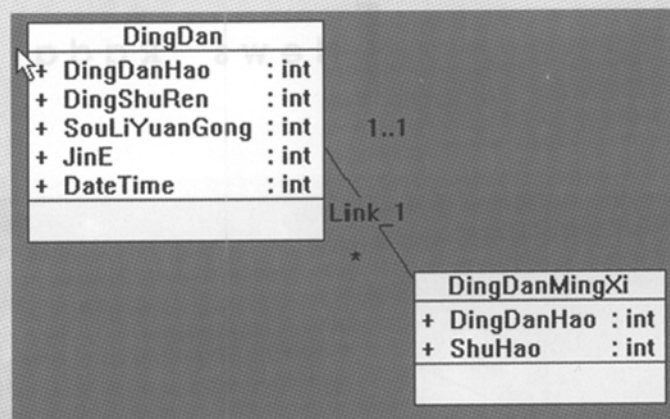


图1

| SouLiYuanGong? | JinE? | DateTime?>

<!ATTLIST DingDan

link_1 IDREFS>

<!ELEMENT DingDanHao (#PCDATA)>

<!ELEMENT ShuHao (#PCDATA)>

<!ELEMENT DingDanMingXi (DingDanHao? | ShuHao?)>

<!ATTLIST DingDanMingXi link_1 IDREF>

由上述生成的 DTD 文件可看出，从 UML 工具中的到 DTD 文件的映射关系为：

类→元素类的属性→元素的子元素
每一个属性→一个#PCDATA 元素
关联 association→DTD 的属性
操作、继承、链接均被忽略。

我们可以对生成的 DTD 文件做些修改，比如对子元素的出现次数的修改，属性或元素的名称的修改。然后可以通过 Powerdesigner 的 Reverse engineer XML 操作把修改过后的 DTD 文件逆映射到 Powerdesigner 的 OOM 图中去。这一次的映射是：

元素→类

元素的子元素→类的属性

DTD 的属性→类之间的关联

通过这种方法，我们就可更灵活、直观、方便、高效地设计出合理的 DTD 文件来。而且在生成 DTD 文件后，可以切换当前的对象语言，利用旧的 OOM 图，在其上调整类的属性的类型，并添加相应的操作和关联，然后就可导出为应用程序的类(例如 Java 语言或 C++) 和相应的后台数据库的物理数据模型。当 DTD 文件表示的逻辑结构有什么变化时，比如添加或删除一个元素时，相应地在建模工具中修改并重新生成，这样就可保证在前台的表现方式(XML 文件)、后台的应用程序(JSP 或 CGI)、和后台的数据库(SQL Server 或 Oracle)之间建立一种一致性。

4 结束语

经过我们的实践，可看出在用 XML 构筑电子商务时，适当地引入 UML 建模技术是一种行之有效的方法，它极大地提高了工作效率，也避免了一些不必要的错误。XML 是一种正日新月异的技术，更新更高效的建模方法仍有待探讨和提高。

参考文献

- 1 《XML 揭秘-入门·应用·精通》，清华大学出版社，Michael Morrison, et al.
- 2 《UML 对象设计与编程》，北京希望电子出版社，刘润东。