对XML Schema的认识

**引言**：可扩展标记语言架构是以可扩展标记语言（标准通用标记语言的子集）为基础的，它用于可替代文档类型定义（外语缩写：DTD）；一份XML schema文件描述了可扩展标记语言文档的结构。XML与软件、硬件和应用程序无关，在数据交换领域有时候可以作为中介来实现某些操作，从而提高开发效率。

**关键词**：XML Schema；XML数据交换

**一、定义**

它的作用是定义一份XML文档的合法组件群，就像文档类型定义（外语缩写：DTD）的作用一样,一份XML Schema定义了：

* 可以出现在文档里的元素；
* 可以出现在文档里的属性；
* 哪些元素是子元素；
* 子元素的顺序；
* 子元素的数量；
* 一个元素应是否能包含文本，或应该是空的；
* 元素和属性的数据类型；
* 元素和属性的默认值和固定值。

**二、优点与劣点**

XML Schema 比 DTD 更强大。其优势包括以下几点：

1、支持数据类型

XML Schema 最重要的能力之一就是对数据类型的支持。通过对数据类型的支持：

* 可更容易地描述允许的文档内容
* 可更容易地验证数据的正确性
* 可更容易地与来自数据库的数据一并工作
* 可更容易地定义数据约束（data facets）
* 可更容易地定义数据模型（或称数据格式）
* 可更容易地在不同的数据类型间转换数据

2、它使用 XML 语法

另一个关于 XML Schema 的重要特性是，它们由 XML 编写。由 XML 编写 XML Schema 有很多好处：

* 不必学习新的语言
* 可使用 XML 编辑器来编辑 Schema 文件
* 可使用 XML 解析器来解析 Schema 文件
* 可通过 XML DOM 来处理 Schema
* 可通过 XSLT 来转换 Schema

3、可保护数据通信

当数据从发送方被发送到接受方时，其要点是双方应有关于内容的相同的"期望值"。

通过 XML Schema，发送方可以用一种接受方能够明白的方式来描述数据。

一种数据，比如 "03-11-2004"，在某些国家被解释为11月3日，而在另一些国家为当作3月11日。

但是一个带有数据类型的 XML 元素，比如：<date type="date">2004-03-11</date>，可确保对内容一致的理解，这是因为 XML 的数据类型 "date" 要求的格式是 "YYYY-MM-DD"。

4、可扩展性

XML Schema 是可扩展的，因为它们由 XML 编写。

通过可扩展的 Schema 定义，您可以：

* 在其他 Schema 中重复使用您的 Schema
* 创建由标准类型衍生而来的您自己的数据类型
* 在相同的文档中引用多重的 Schema

5、可捕获到错误

即使文档的形式良好，仍然不能保证它们不会包含错误，并且这些错误可能会产生严重的后果。

请考虑下面的情况：您订购的了 5 打激光打印机，而不是 5 台。通过 XML Schema，大部分这样的错误会被您的验证软件捕获到。

劣点

虽然被写入XML是一个优势,在某些方面也是一个缺点。W3C XML Schema语言特点是可能非常冗长,而DTD可以简洁且相对容易编辑。

W3C XML Schema没有实现大部分提供的数据元素到文档的DTD能力。

**劣点**

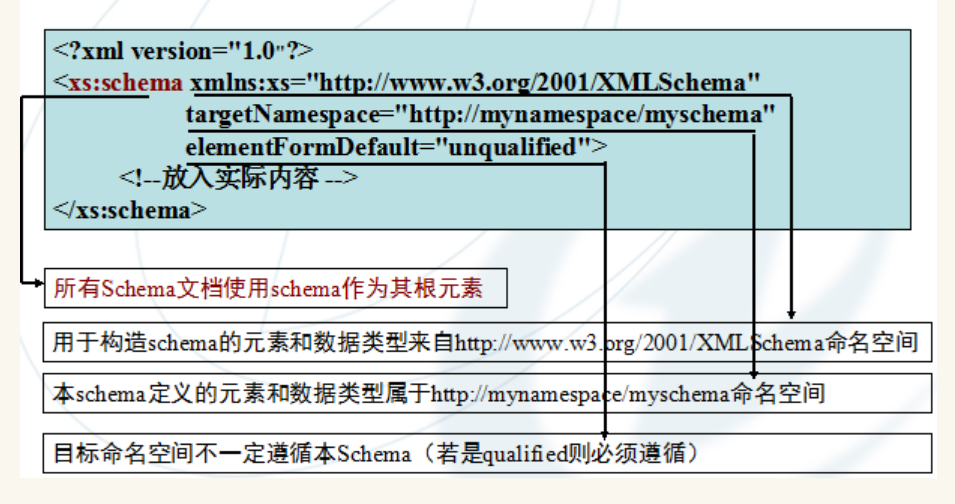
虽然被写入XML是一个优势,在某些方面也是一个缺点。W3C XML Schema语言特点是可能非常冗长,而DTD可以简洁且相对容易编辑。

W3C XML Schema没有实现大部分提供的数据元素到文档的DTD能力。

**三、实际应用**

在现代信息系统中,企业与企业之间出于各种关系和目的,经常需要进行不同数据源之间的数据交换操作。即使在一个企业内部,信息系统升级或数据结构需求等变化,也需要进行数据转换的操作。要实现这些数据转换,仅仅对数据源进行物理层面上的连接是不够的,还需要建立逻辑上的连接,以完成系统中数据结构与数据语义的完整交换。XML已经成为数据交换的标准,而XML的数据模型——XML Schema也就成为一个通用的中间数据模型,其它不同数据模型可通过适配器与其进行交互,最终完成它们之间的数据转换。要完成2个在XML Schema上不同模式结构数据的转换,需要解决两方面的问题:(1)建立两模式上的关联,即确定源模式与目标模式的映射关系;(2)根据模式匹配映射的结果,进行两个数据实例的转换。XML Schema能够清晰有效地定义、解析数据结构,但缺少能够对其中语义进行显示描述的语义元数据,无法让机器自动识别其内含的语义。

**四、Schema文档结构**



一个完整的示例代码：

<xs:element name="purchaseOrder" type="PurchaseOrderType"></xs:element>

<xs:element name="comment" type="xs:string"></xs:element>

<xs:complexType name="PurchaseOrderType">

<!--下面是子元素:依次出现下面子元素-->

<xs:sequence>

<xs:element name="shipTo" type="USAAddress"></xs:element>

<xs:element name="billTo" type="USAAddress"></xs:element>

<xs:element name="items" type="Items"></xs:element>

</xs:sequence>

<xs:attribute name="orderDate" type="xs:date"></xs:attribute>

</xs:complexType>

<xs:complexType name="USAAddress">

<xs:sequence>

<xs:element name="name"></xs:element>

<xs:element name="street"></xs:element>

<xs:element name="city"></xs:element>

<xs:element name="state"></xs:element>

</xs:sequence>

<xs:attribute name="country" type="xs:string" fixed="US"></xs:attribute><!--fixed="US"表示 如果没有 声明此属性，那么默认 就是 US-->

</xs:complexType>

<xs:complexType name="Items">

<xs:sequence>

<xs:element name="productName" type="xs:string"></xs:element>

<xs:element name="quantity">

<xs:simpleType>

<xs:restriction base="xs:positiveInteger">

<xs:maxExclusive value="100"></xs:maxExclusive>

</xs:restriction>

</xs:simpleType>

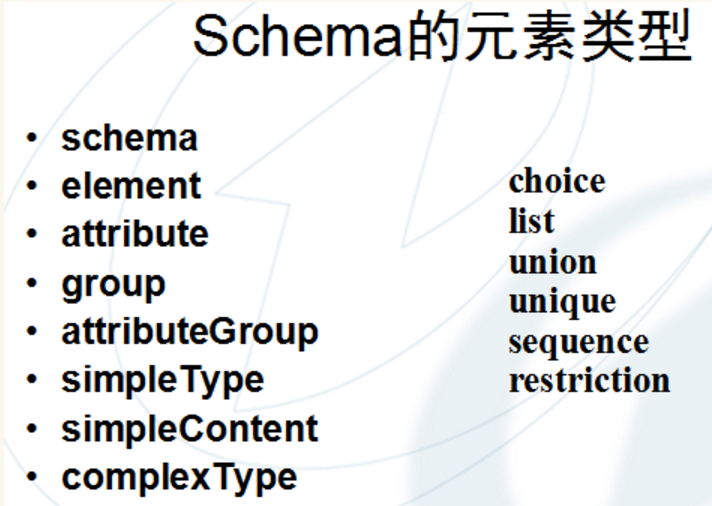
</xs:element>

</xs:sequence>

</xs:complexType>

</xs:schema>

**五、Schema元素类型**



1、element元素

作用：声明一个元素

2、group元素

属性：name、ref

3、attribute元素（为元素声明属性）

属性：name、type、ref

4、attributeGroup元素

作用：将一组属性声明在一起，一边被复杂类型重复使用

属性：name、ref

5、simpleType元素

作用：定义一个简单类型，它决定了元素和属性值的约束和相关信息

属性：name

内容：应用已经存在的简单类型，三种方式

* **r**estrict→限定一个范围
* list→从列表中选择
* union→包含一个值的集合

6、complexType元素

作用：定义一个复合类型，它决定了一组元素和属性值的约束和相关信息

属性：name

（\*）complexType与simpleType区别：

1、simpleType类型的元素中不能包含元素或者属性。

2、当需要声明一个元素的子元素和/或属性时，用complexType；

3、当需要基于内置的基本数据类型定义一个新的数据类型时，用simpleType。

7、simpleContent元素

作用：应用于complexType，对它的内容进行限制和扩展

8、choice元素

作用：允许唯一的一个元素从一个组中被选择

属性：maxOccurs、minOccurs

9、unique元素

作用：定义一个元素或属性值，它必须在特定的范围内

10、sequence元素

作用：按照顺序出现某些元素

参考文献：

【1】百度百科<https://baike.baidu.com/item/XML%20Schema/5587507?fr=aladdin>

【2】中国知网<http://xuewen.cnki.net/CJFD-JSJC200712033.html>

【3】博客园http://www.cnblogs.com/xiaoxiaogogo/p/3601299.html