

文章编号: 1000-5013(2009)06-0646-04

XML 模型到关系模型的映射

洪 欣¹, 陈维斌¹, 蹇崇军²

(1. 华侨大学 计算机科学与技术学院;

2. 华侨大学 机电及自动化学院, 福建 泉州 362021)

摘要: 针对 XML 数据与关系数据在静态数据转换方面存在的问题, 创建一个描述 XML 的 XML model (XML 模型). 使用 XML model 和 ER 模型分别表示 XML 模式和关系模式, 并分析模型间的映射关系, 建立 XML 模式与关系模式的映射框架; 给出从 XML 模式到关系模式的映射算法, 应用该算法可以将 XML 数据转化为关系数据.

关键词: XML; 关系模型; 映射算法; 模型映射

中图分类号: TP 311.132.3

文献标识码: A

XML(可扩展标记语言)的广泛应用,使得如何有效地访问和管理 XML 文档中的数据成为研究的热点. 当前 XML 文档的存储主要分为以下 4 种: (1) 以文件形式存储在文件系统中; (2) 在原生 XML 数据库中存储; (3) 在面向对象数据库中存储; (4) 在关系数据库中存储. 在实际使用过程中, 文件系统无法处理大的 XML 文档, 数据提取和集成存在较大的困难; 原生 XML 数据库则存在数据存取不够高效, 存在安全隐患等问题; 面向对象数据库系统还不足以完成对大型数据库的复杂处理. 关系数据库不仅可以存储海量数据, 而且允许用户并行处理数据, 支持版本控制和不同级别的安全访问机制, 提供可重用信息的共享. 在关系数据库中存储 XML 数据, 已经成为一种常用的策略^[1]. 关系数据与 XML 数据的数据结构存在较大差异, 前者是二维的数据表, 后者是半结构化的数据. 为了实现两者的相互映射, 本文创建了一个 XML model (XML 模型), 并建立 XML 模式与关系模式的映射框架.

1 建立 XML model

由于 XML 数据既包含数据和数据关系, 又包含命令等控制信息, 因此, ER (Entity Relation) 模型和 ODMG (Object Database Management Group) 模型都不适合于描述 XML 数据^[2-3]. XML model 能够完整、直观、准确地描述 XML 数据.

1.1 定义

XML model 将 XML 数据分为 3 种类型并分别采用不同方式表示. 其中, XML model-Control 表示控制信息(包括命令), XML model-Relation 表示关系, XML model-Value 表示数值.

1.1.1 基本规则 在 XML model 中, 虚线指向属性, 实线指向元素.

1.1.2 符号规则 第 1 层是“控制节点(XML model-Control)”, 记录 XML 的注释、CDATA 段、序言和文件类型声明及处理指令, 用圆点线表示. 第 2 层是“关系节点(XML model-Relation)”, 记录 XML 的元素和属性及它们之间的关系, 用箭头线表示, 数字标注表示节点出现次数的下限及上限. 例如: (1, 1) 表示一次(默认值, 可以不写), (1, *) 表示一到无限次, (0, *) 表示零次到无限次. 第 3 层是“数值节点(XML model-Value)”, 记录 XML 的元素和属性的值.

用 XML model 表示的 XML 文件是树状结构. 根节点为 XML 文件名, 中间节点表示元素或属性, 节点名称为元素名或属性名, 叶子节点表示元素或属性的值.

收稿日期: 2009-03-16

通信作者: 洪 欣(1977-), 女, 讲师, 主要从事数据库领域的研究. E-mail: xinhong@hqu.edu.cn.

基金项目: 华侨大学科研基金资助项目(08H-ZR18)

1.2 应用

XML model 可以表示任何符合 XML 语法的数据, 包括 XML 数据、XML 模式, 以及 XML 模式的模式. 以下用 3 个实例分别验证 XML model 的可行性和正确性.

(1) 用 XML model 表示 XML 数据. 其 XML 模式的示例文件如下:

// XML 文件名称: XMLGame.xml	<StarName>里瓦尔多
<? xml version= "1.0" encoding=	</StarName>
"GB 2312" ?>	</Star>
<? xmlstylesheet type= "text/xsl"	</Opponent>
href= "XSLGame.xsl" ?>	<Opponent>
<CnGame xmlns: xsi= "http: // www.	<TeamName ID= "C2" >土耳其
w3.org/2001/XMLSchema	</TeamName>
instance" >	<Place>韩国汉城</Place>
<Opponent>	<Date>2002 年 6 月 13 日</Date>
<TeamName ID= "C1" >巴西	<Time>14:30</Time>
</TeamName>	<Star>
<Place>韩国西归埔</Place>	<StarName >哈坎-苏克
<Date>2002 年 6 月 8 日</Date>	</StarName>
<Time>19:30</Time>	<StarName >奥坎</StarName>
<Star >	</Star>
<StarName>罗纳尔多	</Opponent>
</StarName>	</CnGame>

由于页面限制, 采用分层图来表示该 XML 数据所对应的 XML model, 如图 1 所示.

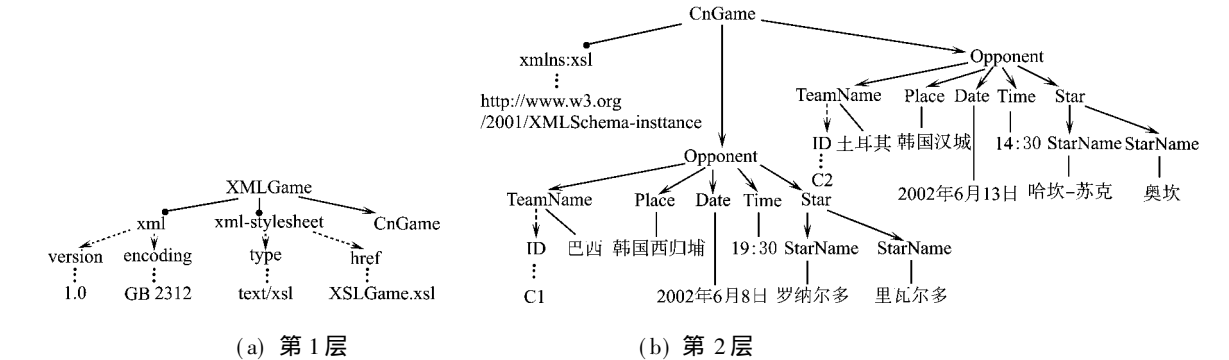


图 1 XMLGame 的 XML model

Fig. 1 The XML model of XMLGame

(2) 用 XML model 表示 XML 模式. XML model 如图 2(a) 所示, XML 模式的示例文件如下:

⋮	<element type= "Star" />
<ElementType name= "Opponent" >	</group>
<attribute type= "ID" />	<element type= "Coach" />
<element type= "TeamName" />	</group>
<group minOccurs= "1" maxOccurs= "1" >	</ElementType>
<group minOccurs= "0"	⋮
maxOccurs= "*" >	

(3) 用 XML model 表示 XML 模式的模式. 由于 XML 模式文件是符合 XML 语法的, 故 XML model 也可用于描述 XML 模式的模式. 图 2(b) 为 XML 模式对应的 XML model, 其定义如下:

<group order= "one"或"seq"或"many"	maxOccurs= "1"或"*">
minOccurs= "0"或"1"	⋮

</group>

1.3 XML model 的特点

由上述 3 个示例的验证表明, XML model 描述 XML 是合适而准确的. 在表示 XML 数据、XML 模式及 XML 模式的模式时, XML model 图形有如下 3 个特点. (1) 控制节点. 在 XML 数据中没有属性的控制节点, 而 XML 模式对应的 XML model 图中存在这样的节点. (2) 数据节点. 与 XML 数据对应的 XML model 的数据节点不存在循环. (3) 数值节点. 在 XML 模式中没有数值节点, 即在与 XML 模式对应的 XML model 图中不会出现.

2 映射分析

关系数据库的建模采用 ER 模型.

ER 模型用实体、联系和属性的概念表示关系数据. 在关系数据库中, ER 模型的实体对应于关系表, 联系对应于表间的关系, 属性对应于字段. XML model 包括 XML-Relation, XML-Control 和 XML-Value, 而 ER 模型不包含控制信息. 因此, 从 XML model 到 ER 模型的映射是有损映射^[3]. XML 模式与关系模式的映射算法如下: 根节点对应于表, 中间节点对应于关系, 叶子节点(包括元素和属性)对应于表的字段, 循环子节点对应于子表. 图 1 经过 XML 模式与关系模式的映射规则转换后的 ER 图, 如图 3 所示.

文档类型定义(DTD)本身不是 XML 格式的文件, 不能定义复杂的数据类型. 因此, 通常采用 XSD(XML 结构定义)或 XDR(XML 数据简化构架定义)来描述 XML 的模式^[4,5]. 由于 XSD 与 XDR 都符合 XML 语法, 可以用 XML model 描述 XSD, XDR 模式. 以 XDR 模式为例, 根据前述的映射示例和算法, 将 XDR 模式映射为关系模式.

(1) 原始的 XDR 模式. XDR 模式描述了一场比赛的 XML 数据的结构: CnGame 节点记录了比赛场次, 在每个场次中通过 Opponent 节点记录比赛信息, 在每个 Opponent 节点中除了描述基本比赛情况 (TeamName, Place, Date, Time), 还在 Star 节点列出参与比赛的明星及教练的名单.

```
<? xml version= " 1.0" encoding= " GB 2312" ?>
<Schema name= " Schema_CnGame"
  xmlns= " urn: Schema-microsoft-com: xml-data"
  xmlns: dt= " urn: Schema-microsoft-com: datatype" >
  <ElementType name= " CnGame" content= " eltOnly" >
    <group minOccurs= " 1" maxOccurs= " * " >
      <element type= " Opponent" />
    </group>
  </ElementType>
  <ElementType name= " Opponent" content= " eltOnly" dt: type= " int" >
    <element type= " TeamName" />
    <element type= " Place" />
    <element type= " Date" />
    <element type= " Time" />
    <group minOccurs= " 0" maxOccurs= " * " >
      <element type= " Star" />
    </group>
  </ElementType>
</Schema>
```

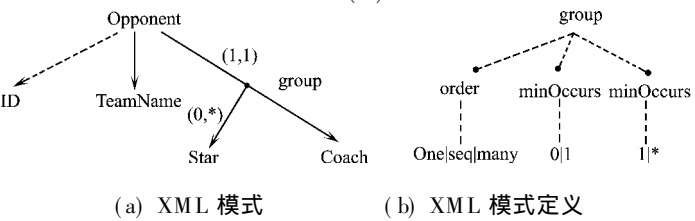


图 2 XML 模式的 XML model

Fig. 2 The XML model of XML Schema

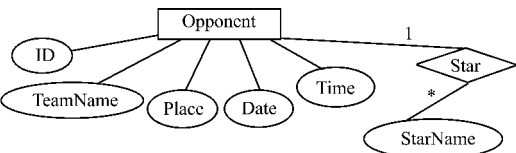


图 3 经过模式映射后的 ER 模型

Fig. 3 The ER model of mapping result

```
<ElementType name= "Star" content= "textOnly" dt: type= "Char" >
  <element type= "StarName" / >
</ElementType>
<ElementType name= "TeamName" content= "textOnly" dt: type= "Char" / >
<ElementType name= "Place" content= "textOnly" / >
<ElementType name= "Date" content= "textOnly" dt: type= "DateT ime" / >
<ElementType name= "Time" content= "textOnly" dt: type= "SmallDateT ime" / >
<ElementType name= " StarName " content= "textOnly" dt: type= "Char" / >
</Schema>
```

(2) 生成的关系模式. 根据 XDR 模式的节点关系, 生成对应的三级关系, 分别是比赛场次表 (CnGame)、比赛信息表(Opponent) 和明星信息表(Star) . 在关系数据库中的映射结果, 如图 4 所示.

(3) XML 数据与关系数据的映射. XML 数据遵循 XML 模式的规则, 同时关系数据遵循关系模式的规则. 因此, 只要按照 XML 模式与关系模式的映射规则填充数据, 就可以实现 XML 数据与关系数据的映射.



图 4 映射生成的关系模式

3 结束语

Fig.4 The relational schema of mapping result

本文建立了一个描述 XML 的模型 XML model, 解决了 XML 缺乏模型表示的问题. 示例的验证表明, XML model 描述 XML 是合适而准确的. 用 XML model 表示 XML 模式, 用 ER 模型表示关系模式, 给出从 XML 模式到关系模式的映射算法, 实现从 XML 数据到关系数据库的映射.

参考文献:

[1] 李占波, 李 娜. XML 数据在关系数据库中的存储[J] . 微计算机信息, 2007, 23(9) : 192- 194.
[2] 王 珊, 萨师煊. 数据库系统概论[J] . 4 版. 北京: 高等教育出版社, 2006.
[3] 洪 欣. 基于 XDR 模式的 XML 文档与关系数据库的映射技术研究[D] . 泉州: 华侨大学, 2004.
[4] 肖继海, 陈俊杰. XML Schema 数据到对象关系数据的转换: 一种基于 DOM- Char 模型的映射算法[J] . 太原理工大学学报, 2006(S2) : 45- 48.
[5] GOLDMAN R, M CHUGH J, WIDOM J. Form semistructured data to XML: Migrating the lore data model and query language[C] // ACM SIGMOD Workshop on the Web and Databases Philadelphia: [s. n.], 1999: 25- 30.

Mapping from XML Model to Relational Model

HONG Xin¹, CHEN Weir bin¹, JIAN Chong-jun²

(1. College of Computer Science and Technology;
2. College of Mechanical Engineering and Automation, Huaqiao University, Quanzhou 362021, China)

Abstract: To solve the problem of static data transformation from XML (extensible markup language) data to relational data, a XML model is established to describe XML. The XML pattern and relation pattern are expressed respectively by XML model and entity- relation (ER) model and to analyze the mapping relation of the two models. The mapping framework from XML pattern to relation pattern is established. The mapping algorithm from XML pattern to relation pattern is presented to transform the XML data to relational data.

Keywords: XML; relational model; mapping algorithm; model mapping