

# EER模型中超类/子类到XML模式的转换研究\*

唐远翔, 刘 益

(宜宾学院 计算机与信息工程学院, 四川 宜宾 644007)

**【摘要】**针对目前所使用的增强型ER模型, 本文介绍了对该模型中的超类/子类, 特化和概化, 以及超类型/子类型联系中的约束的无损语义转换方法。实践证明该转换方法具有很好的实用性, 能对后续的算法设计起到很好的作用。

**【关键词】**关系模式; XML schema; 模式转换; EER; DDT

**【中图分类号】**TP311.132.3 **【文献标识码】**A **【文章编号】**1673-1891(2012)02-0075-03

## 引言

XML是Extensible Markup Language(可扩展的置标语言)的缩写, 是W3C组织于1998年2月发布的标准。XML源于SGML(Standard for General Markup Language, 简称SGML), 是SGML的一个简化版本。20世纪90年代中期以来, XML以其丰富的语义和灵活的结构逐渐成为网络环境下数据交换的标准<sup>[1]</sup>。现今由于关系数据库的优势, 其被广泛应用, 如何把关系数据库中的数据转换成XML以实现在网络上的传输已成为研究热点。本文从EER模型出发给出了一种实现无损语义转换的方法。

## 1 EER的特点

实体联系模型(Entity-Relationship model)是P. P.S.Chen于1976年提出的, 它适用于大多数一般的商业问题, 至今仍然得到广泛的研究和应用。但现今的商业环境及商业关系的变化和发展导致了商业数据之间关系的复杂性进一步加剧, 为了应对这些变化, 研究人员对ER模型进行了改进, 提出了增强型实体-联系(Enhanced Entity-Relationship, EER)模型。该模型更能准确地描述当前商业环境中的复杂数据。EER模型是在ER模型基础上增加语义表达形成的。主要体现在对ER模型进行扩充: ①D.R.Howe为了增加ER模型中实体间联系的语义, 引进了“属性类”的概念。②T.J.Teorey和T.P.Fry等人采用不同的图记号来表示ER图的语义, 并且在ER图中还增加了进行“数据抽象”的“归纳”层次图示法<sup>[2,3]</sup>。

### 1.1 超类与子类

EER模型中包含的最重要的建模产品是超类型/子类型。它简化了对一般实体型(称为超类型)的建模, 然后将其划分成一些具体的实体类型(称为子类型)。每个子类型继承超类型的属性, 并且可以增加自己的特殊属性。一个子类型是组织中

有效的实体型中的实体的子集, 它享有公共属性或关系, 同时又与其他子集相区别。引入超类/子类基于如下两个原因: ①便于区分某些具有特殊属性的实体。②特殊关联的实现。

超类与子类有两个性质: ①子类与超类之间具有继承性。②继承性是通过子类实体和超类实体有相同的实体标识符实现<sup>[2,3]</sup>。超类/子类的概念引入ER模型能带来两个好处。第一, 避免了对相似概念进行重复的描述, 使ER图具有更好的可读性。第二, 为设计添加了更多的语义信息, 能使数据模型传递更多的语义内容。

### 1.2 概化与特化

概化(Generalization)/特化(Specialization)是EER模型的主要概念, 它与超类/子类及其继承有关。概化是一个在比较具体的实体类型集合中抽象出一个比较概括的实体类型的过程, 是自底向上的。与概化相反, 特化是一个自顶向下的过程, 是定义一个或多个属于超类型的子类型, 并建立超类型/子类型之间联系的过程<sup>[2-4]</sup>。

### 1.3 超类型与子类型联系中的约束

在超类型与子类型联系中的约束有两个重要的类型约束。分别是完全性约束和非重叠性约束。

完全性约束是指一个属于某一超类型的实例是否必须至少是一个子类型的实例。非重叠性约束是指一个属于某一超类型的实例是否可以同时属于两个或两个以上的子类型。

## 2 XML模式

XML(Extensible Markup Language)于1998年2月被引入软件工业界, 现已成为Internet上用来表示和交换数据的标准, XML标记能够自我描述, 可以定义新的标记和属性。XML模式是指用来描述XML结构、约束等因素的语言, 例如XML Schema、XML DTD、XDR、SOX等等。XML格式则是XML文

收稿日期: 2012-03-22

\*基金项目: 宜宾学院青年基金(NO QJ05-03)资助项目。

作者简介: 唐远翔(1979-), 男, 四川越西人, 讲师, 主要研究方向: 逻辑推理、数据库设计、智能信息处理。

档本身所具有的格式。从模式的描述语言来说, XML Schema 和 XML DTD 都属于语法模式。与概念模式不同, 语法模式在对同一事物描述时, 可以采用不同的语法, 例如在对关系模式描述时, 无论是使用 XML Schema, 还是 XML DTD, 都既可以用元素, 也可以用属性来描述关系模式的列<sup>[9]</sup>。

本文将着重考虑 EER 模式到 DTD 之间的转换。

### 3 转换方法

#### 3.1 超类/子类到 XML

超类/子类定义了子实体和超实体之间的关系。在模式映射到 DTD 中, 把每一个子实体映射成 DTD 中的子元素, 而超实体映射成父元素, 每个父元素可以有零个或多个子元素, 如图 1 所示。



图1 模式转换1

关系模式	DTD
A(A1, A2) B(*A1, B2)	<!ELEMENT A(B)?>
	<!ATTLIST A
	A1 CDATA #REQUIRED
	A2 CDATA #REQUIRED>
	<!ELEMENT B EMPTY>
	<!ATTLIST B
	B1 CDATA #REQUIRED>

#### 3.2 概化、特化的模式转换

概化描述了实体之间的联系, 将多个实体所共有的“一般化”属性提取出来形成一个新的实体, 也就是这些具有共同特性的实体的超类或子类。在 DTD 中我们把这个“一般化”的超类提取出来映射成一个元素, 而子类就映射成相应的子元素, 如图 2、图 3 所示。

##### 3.2.1 非重叠模式转换

在模式映射到 DTD 中, 把每一个子实体映射成 DTD 中的子元素, 而超实体映射成父元素, 每个父元素只能允许拥有一个子元素。

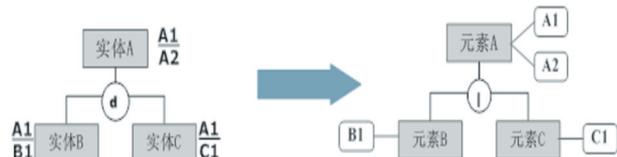


图2 模式转换2

关系模式	DTD
A(A1, A2) B(*A1, B2) C(*A1, C1)	<!ELEMENT A(B C)>
	<!ATTLIST A

```
A1 CDATA #REQUIRED
A2 CDATA #REQUIRED>
<!ELEMENT B EMPTY>
<!ATTLIST B
B1 CDATA #REQUIRED>
<!ELEMENT C EMPTY>
<!ATTLIST C
C1 CDATA #REQUIRED>
```

##### 3.2.2 重叠模式转换

在模式映射到 DTD 中, 把每一个子实体映射成 DTD 中的子元素, 而超实体映射成父元素, 每个父元素允许拥有零个或多个子元素。

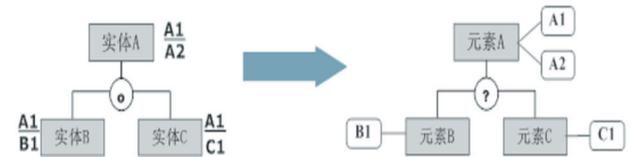


图3 模式转换3

关系模式	DTD
A(A1, A2) B(*A1, B2) C(*A1, C1)	<!ELEMENT A(B?, C?)>
	<!ATTLIST A
	A1 CDATA #REQUIRED
	A2 CDATA #REQUIRED>
	<!ELEMENT B EMPTY>
	<!ATTLIST B
	B1 CDATA #REQUIRED>
	<!ELEMENT C EMPTY>
	<!ATTLIST C
	C1 CDATA #REQUIRED>

##### 3.2.3 分类转换

子类是超类间的并集的子集, 也就是子类中的数据能且只能出现在一个超类之中。在模式映射到 DTD 中, 把超类映射成元素, 超类间共同的子类映射成同层次的元素。元素间通过 ID 与 IDREF 实现相应的联系。

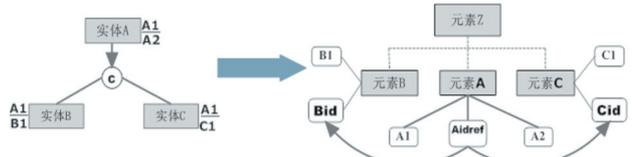


图4 模式转换4

关系模式	DTD
B(A1, A2) C(*A1, B2) A(*A1, C1)	<!ELEMENT B EMPTY>
	<!ATTLIST B
	B1 CDATA #REQUIRED
	id ID #REQUIRED>
	<!ELEMENT A EMPTY>

<!ATTLIST A	C1 CDATA #REQUIRED
A1 CDATA #REQUIRED	id ID #REQUIRED>
A2 CDATA #REQUIRED	
idref IDREF #REQUIRED>	
<!ELEMENT C EMPTY>	
<!ATTLIST C	

#### 4 结论

根据上述方法,能实现EER模型到XML模式的转换,在转换过程中利用DTD的特点保持了EER中语义约束,达到了无损语义转换。

#### 注释及参考文献:

- [1]王仲,陈晓鸥.基于XML的数据交换与存取技术研究[J].计算机工程与应用,2001(24):108-111.
- [2]吴扬扬,陈锻生.一个基于XML的Web信息源集成方案[J].计算机工程与应用,2001(10):34-37.
- [3]吴爱华,刘小玲,王洪,等.基于混合映射的XML数据的关系存储和查询[J].郑州大学学报:理学版,2007,39(2):157-160.
- [4]吴淑雷,雷景生,陈焕东.基于XML的异构数据交换技术[J].广西师范大学学报(自然科学版),2007,25(2):78-81.
- [5]王志平.基于XML的异质多数据库集成系统的设计与实现[J].河南大学学报(自然科学版),2007,37(5):530-532.
- [6]Mark Birbeck.XML高级编程[M].北京:机械工业出版社,2002:277-346,613-624.

## Research on the Transformation of the EER Model to XML Schema

TANG Yuan-xiang, LIU Yi

(School of Computer and Information Engineering, Yibin University, Yibin, Sichuan 644007)

**Abstract:** Aiming at the existing enhanced ER mode, this paper provides a new transforming method without losing any semantics between superclass and subclass, and the introduction of superclass, subclass, specialization, generalization of the mode is given. This method is effective in practice, which plays a very good effect on subsequent algorithm.

**Key words:** Relational schema; XML schema; Mode conversion; EER; DDT