

# 含有效时间时态关系数据库到 XML 映射方法的研究<sup>\*</sup>

刘 健 马宗民 严 丽

(东北大学信息科学与工程学院 沈阳 110004)

**摘 要** 在讨论现有时态 XML 数据模型的基础上,引入有效时间 XML 表示模型,进而研究含有效时间时态关系数据库到时态 XML 文档的映射问题。给出时态关系数据库到时态 XML 文档映射的一般性映射规则,重点研究时态关系数据库一些复杂映射问题,包括类层次、联系集、弱实体集的映射转换问题。

**关键词** XML, 时态关系数据库, 时态 XML, 映射

## Formal Mapping of Temporal Relational Databases with Valid Times into Temporal XML

LIU Jian MA Zong-min YAN Li

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China)

**Abstract** The valid time is introduced into XML document based on the existing temporal XML data models. The mapping from temporal relational databases with valid time into temporal XML document is investigated in the paper. In addition to some generic mapping rules, some complex mappings, including inheritance hierarchies, relationship set, and weak entity set, are especially studied and proposed.

**Keywords** XML, Temporal relational databases, Temporal XML, Mapping

## 1 引言

人类生活在随时间动态进化的世界里,时间是所有现实世界现象的属性。随着 Internet 技术的迅速发展,新一代基于 Web 信息系统对时态信息处理的需求变得越来越迫切。用于时态信息表示与处理的时态数据库研究始于 20 世纪 70 年代,经过几十年的发展,其在理论研究和应用开发方面都取得了巨大的进步。但是时态数据运算体系目前还不够成熟,还没有统一完整的国际标准,导致了时态数据库模型过多,因而也就造成了应用中的混乱。现有的时态数据库模型和数据管理方法已经不能满足 Web 时代高效现代商业数据管理的需求<sup>[2]</sup>,XML<sup>[1]</sup>的出现为这些问题的解决提供了新的技术支持。

作为 Internet 的下一代标记语言,XML 在电子商务等 Web 应用中发挥着越来越重要的作用。利用 XML 可以很自然地对时态信息进行描述<sup>[9,10]</sup>和查询<sup>[6,7]</sup>,用时态 XML 数据模型处理有效时间、事务时间等相关时态问题<sup>[3-5]</sup>,是对传统数据模型的突破。时态 XML 的研究最近几年受到国内外研究者的重视,发表了一些重要的研究成果<sup>[8,14,18,19]</sup>。这些研究工作主要集中在对时态 XML 模型的表示和时态查询机制的研究,但对于如何把现有的时态数据库资源映射到时态 XML 的研究却非常有限,而传统关系数据库到 XML 的映射<sup>[15,16,20]</sup>是一个被广泛研究的课题。时态关系数据库到时态 XML 的映射,可为时态数据库信息的 Web 发布与共享、实现基于 Web 时态信息的管理奠定坚实的基础。

本文将在讨论现有时态 XML 数据模型的基础上,引入有效时间 XML 表示模型<sup>[6,7,15]</sup>,进而研究含有效时间时态关系数据库到时态 XML 文档的映射问题。文章将给出时态关系数据库到时态 XML 文档映射的一般性映射规则,并重点

讨论时态关系数据库中一些复杂的映射问题,包括类层次、联系集、弱实体集的映射转换问题。

本文组织如下:在第 2 节给出时态 XML 文档的相关知识之后,将讨论时态 XML 文档中有效时间的表示问题,第 4 节将具体研究时态关系数据库和时态 XML 文档之间的映射转换问题,最后总结全文并对以后工作进行展望。

## 2 相关知识

### 2.1 XML 相关知识

XML 是一套定义语义标记的规则,这些标记将文档分成许多部件并对这些部件加以标识。它是元标记语言,即定义了用于定义其它与特定领域有关的、语义的、结构化的标记语言的句法语言。XML 文档分两类:以数据为中心和以文档为中心。以数据为中心的文档特点是结构相当规范、数据颗粒度好。其中同层次元素和 PCDATA 的出现顺序并不重要。以文档为中心的文档的特点是结构不够规范、数据颗粒度大以及含有大量的混合内容,其中相同层次的元素和 PCDATA 的出现顺序是非常重要的。

XML 可以由使用者自由定义标记,并通过元素之间的嵌套包含来体现层次关系。它主要有 3 个要素:DTD 或 Schema<sup>[17]</sup>,XSL,XLL。其中 DTD 或 Schema 规定 XML 文档的逻辑结构,定义 XML 文档中的元素、属性以及元素之间的关系,它可以分析校验 XML 文档的有效性。XSL 用于规定 XML 文档的显示方式,它能够在客户端 Web 浏览器改变文档的显示形式,而不需要在服务器端进行通讯交互。XLL 主要用于进一步扩展目前 Web 上已有的简单链接。XML 的使用者不仅可以定义文档中元素的词汇表,还可以指定元素间的关系,将文档中的元素进行结构化,形成以逻辑内容为基础的标记结构。这种结构化、良构的 XML 文档为信息组

<sup>\*</sup>教育部新世纪优秀人才支持计划(NCET20520288)。刘 健 硕士研究生;马宗民 教授,博士生导师;严 丽 副教授。

织提供了强有力的支持。由于 XML 是一种公共的、普遍适用的数据格式,人们可以使用任何一种理解 XML 的工具来操作数据,而不受到数据格式的限制,因此很适合作为不同系统之间交换数据的标准。

## 2.2 时态关系数据库相关知识

时态信息处理已经成为新一代数据库与信息系统的核心技术,时态数据一般保存在数据库系统中。记录和管理时态信息的数据库,有别于传统数据库,记录中的元组属性带有时变的特性,这就导致其数据模式与传统关系数据库模式差别很大,由此产生了数据库的一些基本概念、理论、性质方面的改变,这种反映时态信息的数据库就是时态数据库<sup>[11,12]</sup>。

在时态数据库中存在两种特殊的时间元素 Now, UC。Now 是一个时间变元,其有效值依赖当前时间。在有效时间中包含 Now 意味着该事实有效的期限,在传统 SQL 数据库中记录有效期的语义一般在应用程序中执行,而在时态数据库中的语义直接架构在数据模型和查询语言中。UC 主要是用于事务时间中的时间变元,表示直到该元素改变的时间,在应用于事务中,它比 Now 有更精确的解释语义。

有效时间是指一个对象在现实世界中发生并保持的那段时间。有效时间可以是单一时间点、单一的时间区间或者是时间点、时间区间的有限集合,也可以是整个时间域。有效时间可以被更新,可以指过去、现在和未来。事务时间对应于现有事务或现有数据库状态变迁的历史。它是应用独立的,用户不可修改事务时间。事务时间不能晚于现在时间,因为它反映着数据库实际操作的时间,不能指未来。

## 3 时态 XML 数据模型

时态 XML 模型为每个实体增加了一个生命期,支持属性统一分组,用时间属性很清晰地表示时态数据信息。假设有四个实体集 students, courses, cars 和 teachers, 实体集 students 中有属性 sname, dept, sname 为主键; 实体集 courses 中有属性 cname, credit, cname 为主键; 实体集 cars 中有属性 dname, origin, dname 为主键; 实体集 teachers 中有属性 tname, title, salary, tname 为主键。以上四个实体集可以用类似的 XML 文档表示, 简便起见, 我们只研究实体集 teachers。假设实体集 teachers 有如下时间属性:

$tname(e) = \{[1995-01-01, Now] \rightarrow Lyot\}$   $title(e) = \{[1995-01-01, 1997-12-31] \rightarrow AsProfessor, [1998-01-01, Now] \rightarrow Professor\}$   $salary(e) = \{[1995-01-01, 1997-12-31] \rightarrow 5000 [1998-01-01, 1999-12-31] \rightarrow 6000 [2000-01-01, Now] \rightarrow 9000\}$  这里属性值都带有一个有效时间生命期。表 1 描述了实体集 teachers 的记录信息, 每个属性都添加了一个时间戳, 用这个时间戳来表示有效时间。这并不是时态分组描述, 图 1 是相应的 XML 文档表示, 这里每个子元素都添加了 vstart 和 vend 两个时间属性, 用来表示有效时间。

表 1 包含有效时间的 teachers 记录信息

tname	title	salary	start	End
Lyot	AssProfessor	5000	1995-01-01	1997-12-31
Lyot	Professor	6000	1998-01-01	1999-12-31
Lyot	Professor	9000	2000-01-01	Now

在包含有效时间的记录信息中, 当满足 i) 记录信息属性值相等, 并且 ii) 记录信息的时间间隔相同或是可以合并时, 两条记录信息就可以合并。表 2 是实体集 teachers 的时态分

组描述。

```

<teachers vstart="1995-01-01" vend="now" >
  <teacher vstart="1995-01-01" vend="now">
    <tname vstart="1995-01-01" vend="now">Lyot</tname>
    <title vstart="1995-01-01" vend="1997-12-31">AssProfessor</title>
    <title vstart="1998-01-01" vend="now">Professor</title>
    <salary vstart="1995-01-01" vend="1997-12-31">5000</salary>
    <salary vstart="1998-01-01" vend="1999-12-31">6000</salary>
    <salary vstart="2000-01-01" vend="now">9000</salary>
  </teacher>
</teachers>

```

图 1 teachers 有效时间记录的 XML 表示

时态分组描述能减少元组信息的冗余、节约存贮空间, 以及提高查询等操作效率。

表 2 时态分组的 teachers 记录信息

tname	title	salary
1995-01-01	1995-01-01	1995-01-01
	AssProfessor	5000
Lyot	1997-12-31	1997-12-31
	1998-01-01	1998-01-01
Now	6000	1999-12-31
	Professor	2000-01-01
	Now	9000
	Now	Now

## 4 时态关系数据库到时态 XML 文档的映射

XML 能很好地描述数据, 具有自描述和可扩展的特性, 并且与平台无关, 因此可以作为一种传递数据的交换标准。

对于包含有效时间的时态关系数据库中的一个关系, 我们可以把有效时间看成 XML 文档中的一个元素的属性, 用一对时间戳属性来表示有效时间。一般来讲我们可以使用如下的映射规则:

(1) 对于时态关系数据库中的关系表, 直接映射生成一个唯一的 XML 根元素, 根元素的起始时间属性为所有子元素起始时间属性的最小值; 终止时间属性为所有元素终止时间属性的最大值, 一般可设定为 Now。

(2) 在根元素内部, 映射生成复杂的 XML 文档子元素, 用来表示时态关系数据库中的记录信息。起始时间和终止时间属性的取值方法与根元素类似。

(3) 对于时态关系表中记录信息为单值的属性, 直接映射到 XML 文档的子元素中, 并为子元素添加 vstart 和 vend 时间属性。

(4) 对于时态关系表中记录信息为多值属性, 当满足属性值相等, 并且属性值的时间间隔相同或可以合并时, 我们就把两条记录信息的时间属性合并成为 XML 文档中的一个子元素属性; 当属性值不相等时, 则用其相应的时间属性标示 XML 文档中的子元素。

以表 1 为例, 时态关系表首先映射生成了一个唯一的根元素 teachers, 然后在其内部生成复杂子元素 teacher, 起始时间属性取定为 "1995-01-01", 而终止时间属性取值为 "Now"。其中 tname 是一个单值属性, 直接把它映射到 XML 文档的一个子元素:

```

<tname vstart="1995-01-01" vend="now">Lyot</tname>

```

而 title 是一个多值属性, 其中 Professor 满足属性值相等, 并

且时间间隔可以合并,我们把它映射到 XML 文档的一个子元素:

```
<title vstart="1998-01-01" vend="now">Professor</title>
```

对于 AssProfessor,它和合并后的 Professor 不满足属性值相等这一条件,所以把它们分开映射成为 XML 文档的两个子元素:

```
<title vstart="1995-01-01" vend="1997-12-31">AssProfessor</title>
<title vstart="1998-01-01" vend="now">Professor</title>
```

每个 XML 子元素都添加了 vstart 和 vend 时间属性,用来表示有效时间,这样就实现了从时态关系数据库到时态 XML 文档的一般性映射转换。

上面给出的映射规则,讨论的是单个简单关系表的映射,而单个关系表以及不同的关系表之间可能存在一些复杂的情况,例如类层次关系、联系集、弱实体集等。下面我们重点讨论类层次关系、联系集以及弱实体集等复杂情况时态关系数据库到 XML 映射转换问题。

### 4.1 类层次的转换

从语义上来讲,子类中的每一个实体同时也是超类中的一个实体,子类中实体属性必须包含所有的超类实体集的属性,即子类将继承超类的属性。表 3、表 4 是类层次关系中的 2 个时态关系,其中 UGstudents 和 Gstudents 是 students 的 2 个子类,students 是 UGstudents 和 Gstudents 的超类,子类继承了超类中的所有属性,同时又具有自己的特殊属性,它们之间的类层次关系用图 2 中的一个特殊三角箭头表示。

表 3 包含有效时间的 UGstudents 记录信息

Sname	dorm	start	end
Lytot	D2	2005-03-01	2007-03-31
Lytot	D9	2007-04-01	Now

表 4 包含有效时间的 Gstudents 记录信息

Sname	title	start	end
Vincent	Master	2004-09-01	2007-02-28
Vincent	Doctor	2007-03-01	Now

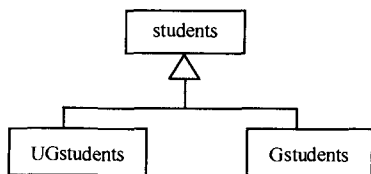


图 2 超类\子类的继承关系

```
<UGstudents vstart="2005-03-01" vend="now">
  <UGstudent vstart="2005-03-01" vend="now">
    <sname vstart="2005-03-01" vend="now">Lytot</sname>
    <dorm vstart="2005-03-01" vend="2007-03-31">D2</dorm>
    <dorm vstart="2007-04-01" vend="now">D9</dorm>
  </UGstudent>
</UGstudents>
<Gstudents vstart="2004-09-01" vend="now">
  <Gstudent vstart="2004-09-01" vend="now">
    <sname vstart="2004-09-01" vend="now">Vincent</sname>
    <title vstart="2004-09-01" vend="2007-02-28">Master</title>
    <title vstart="2007-03-01" vend="now">Doctor</title>
  </Gstudent>
</Gstudents>
```

图 3 包含有效时间的子类信息的 XML 表示

当把图 2 中 3 个具有类层次关系的时态关系映射成 3 个

XML 文档时,时态关系 students 的映射比较简单,可以用前面我们给出的一般映射规则实现。其 XML 文档形式与时态关系 teachers 的类似,因此这里就不给出时态关系 students 的具体实例以及 students 的时态 XML 文档。时态关系 UGstudents 含有属性 sname 和 dorm,其中 sname 是主关键字,同时也是外键。时态关系 Gstudents 含有属性 sname 和 title, sname 是主关键字,又是外键。图 3 是转换后时态关系 UGstudents 和 Gstudents 所对应的时态 XML 表示形式。

### 4.2 联系集的转换

现实世界的事物之间是有联系的。联系可以描述实体间关系,也可以描述实体内部的关系。表 5 是联系集 choosing 的 1 个时态关系,表示多对多联系,这里我们仍用 start, end 表示有效时间。图 4 反映了时态关系 students 和时态关系 courses 之间的多对多联系,用 choosing 表示。

表 5 包含有效时间的 choosing 记录信息

sname	tname	grade	start	end
Lytot	C	95	2006-03-01	2006-05-31
Lytot	C++	89	2006-06-01	Now

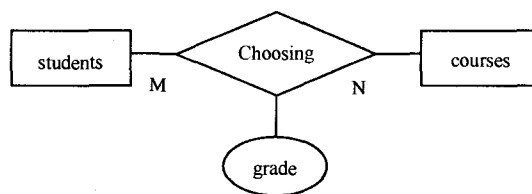


图 4 实体间的联系

当把图 3 中构成联系集关系的 3 个时态关系映射成 3 个 XML 文档时,时态关系 students 和 courses 可以用前面我们给出的一般映射规则实现到 XML 文档的映射。这里我们省去时态关系 students 和 courses 的具体实例以及相应的时态 XML 文档,只关注时态关系 choosing 到时态 XML 的映射。我们把与时态关系 choosing 相关联的 2 个时态关系 students 和 courses 的主关键字和外键分别当作 choosing 的外键和主关键字,把联系集本身的属性作为普通属性并加以时间限制,这样就生成了能描述多对多联系的 XML 文档。图 5 是转换后时态关系 choosing 多对多联系集所对应的时态 XML 文档。

```
<relationships vstart="2006-03-01" vend="now">
  <choose vstart="2006-03-01" vend="now">
    <sname vstart="2006-03-01" vend="now">Lytot</sname>
    <tname vstart="2006-03-01" vend="2006-05-31">C</tname>
    <tname vstart="2006-06-01" vend="now">C++</tname>
    <grade vstart="2006-03-01" vend="2006-05-31">95</grade>
    <grade vstart="2006-06-01" vend="now">89</grade>
  </choose>
</relationships>
```

图 5 包含有效时间的联系集的 XML 表示

### 4.3 弱实体集的转换

弱实体是现实世界的一种特殊的联系,它是依赖其它实体存在而存在的实体。被依赖实体与弱实体之间是一对多的联系,并且弱实体要求全部参与联系。例如,教师需要定期给车买保险,表 6 是其具体花费情况的时态关系。它与被依赖(时态)关系 teachers 和弱实体 cars 之间的联系用图 6 表示,

(下转第 250 页)

**结束语** 本文研究了基于混沌特性的小波数字水印技术。先利用混沌特性对水印快速加密,将加密后的水印信息嵌入到相对重要的小波域系数中。仿真实验表明,本文提出的数字水印技术具有良好的隐蔽性、稳定性和鲁棒性,嵌入水印的能量自适应于图像本身,水印检测时不需要原始载体和原始水印图像,是一种比较理想的变换域盲水印算法,能够满足特定领域的应用需求。

系统对水印信息采用调整区间的做法,在一定程度上减少了水印嵌入对原始图像在视觉上的影响。调整的区间很好地适应于原始图像的嵌入区间,使嵌入前后的小波系数差别变小,载体图像在嵌入前后在视觉上没有变化。

算法采用设定嵌入区间的办法使嵌入前后的小波系数相关联,这种方法的最大好处是提取水印时不依赖于原始图像和水印,使提取出的水印在所有权验证时更加具有说服力,适应了数字水印的发展趋势。

(上接第 242 页)

其中虚方框表示弱实体集,带有虚线的箭头指向弱实体集。

表 6 包含有效时间的 policy 记录信息

Tname	dname	origin	cost	start	end
Lyot	volvo	Sweden	700	2005-03-01	2006-03-01
Lyot	BWM	Germany	900	2006-03-02	Now

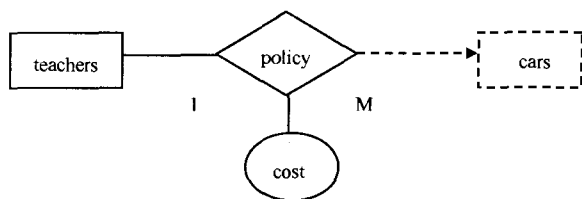


图 6 弱实体集关系

在时态 XML 文档中,我们把弱实体集和联系集合并形成一个 XML 文档,表示弱实体的概念,同时把参照实体集的主关键字作为外键,与弱实体集的主关键字共同构成该实体集的主关键字。图 7 是转换后时态关系 policy 所对应的时态 XML 文档。

```

<relationships vstart="2005-03-01" vend="now">
  <policy vstart="2005-03-01" vend="now">
    <tname vstart="2005-03-01" vend="now">Lyot</tname>
    <dname vstart="2005-03-01" vend="2006-03-01">volvo</dname>
    <dname vstart="2006-03-02" vend="now">BWM</dname>
    <origin vstart="2005-03-01" vend="2006-03-01">Sweden</origin>
    <origin vstart="2006-03-02" vend="now">Germany</origin>
    <cost vstart="2005-03-01" vend="2006-03-01">700</cost>
    <cost vstart="2006-03-02" vend="now">900</cost>
  </policy>
</relationships>

```

图 7 包含有效时间的弱实体集的 XML 表示

**结束语** 随着 Web 的广泛应用,XML 正发挥着越来越重要的作用,如何通过 XML 更有效地表示和存储时态信息,已经成为人们研究的热点。本文讨论了时态 XML 文档中的有效时间表示问题,该模型既可以处理以文档为中心的应用,也可以处理以数据为中心的应用。为了能对大量的时态数据进行有效管理,我们重点研究了时态关系数据库和时态 XML 文档之间的映射问题,总结出了从时态数据库到时态 XML 文档的转换方法,该方法丰富了基于 XML 的时态运算体系。

## 参考文献

- [1] Turner L F. Digital data security system. Patent IPN WO 89/08915, 1989
- [2] van R G S, Tirkel A Z, Osborne D F. A digital watermark. Int Confon Image Processing, 1994 (2): 86-90
- [3] Brassil J, Low S, Maxemchuk N, et al. Electronic marking and identification techniques to discourage document copying//Proc. of Infocom. 94, 1994
- [4] Bender W, Gruhl D, Morimoto N, et al. Technique for data hiding. IBM System Journal, 1996, 35(4)
- [5] Cox I. Secure spread spectrum watermarking for multimedia. IEEE Trans on Image processing, 1997, 6(12): 1673-1678
- [6] Podilchuk C I, Zeng W. Image-adaptive watermarking using visual models [J]. IEEE Journal on Select Areas Commun., 1998, 16(4): 525-539
- [7] 张小飞. 基于混沌特性的小波数字水印算法研究. 硕士学位论文论文. 2007, 5: 26-29
- [8] 丁玮, 闫伟齐, 齐东旭. 基于生命游戏的数字图象置乱与数字水印技术. 北方工业大学学报, 2000, 12(1): 1-4
- [9] Gonzalez R C, Woods R E. Digital Image Processing Second Edition. Publishing House of Electronics Industry, 2003

今后的研究工作重点将放在时态 XML 存储空间和查询性能的优化技术问题上,以及研究时态 XML 文档的索引和更新技术,以便建立一套基于 XML 的完备时态操作运算体系。

## 参考文献

- [1] World Wide Web Consortium. Extensible Markup Language (XML) 1.0. http://www.w3.org/TR/REC-xml. W3C Recommendation. February 1998
- [2] Ozsoyoglu G, Snodgrass R T. Temporal and Real-time Databases: A Survey. IEEE Transactions on Knowledge and Data Engineering, 1995, 7(4): 513-532
- [3] Jensen C S, et al. A Consensus Glossary of Temporal Database Concepts. ACM SIGMOD Record, 1994, 23(1): 52-64
- [4] Allen J F. Towards a General Model and Action and Time. Artificial Intelligence, 1984, 23(2): 123-154
- [5] Kabanza F, Stevenne J-M, Wolper P. Handling Infinite Temporal Data. Journal of Computer and System Sciences, 1995(51): 3-17
- [6] Wang F, Zaniolo C. Temporal Queries in XML Document Archives and Web Warehouses. In TIME-ICTL, 2003
- [7] Wang F, Zaniolo C. Publishing and Querying the Histories of Archived Relational Databases in XML. In WISE, 2003
- [8] Amagasa T, Yoshikawa M, Uemura S. A Data Model for Temporal XML Documents. In DEXA, 2000
- [9] Chawathe S S, Abiteboul S, Widom J. Representing and querying changes in semistructured data//Proceedings of the Fourteenth International Conference on Data Engineering, 1998
- [10] Chawathe S S, Abiteboul S, Widom J. Managing historical semistructured data. TAPOS, 1999, 5(3)
- [11] Gadia S K. A Homogeneous Model and Query Language for Temporal Databases. ACM TODS, 1988, 13(4): 418-448
- [12] Snodgrass R T, et al. TSQL2 Language Specification. ACM SIGMOD Record, 1994, 23(1): 65-86
- [13] Manukyan M G. Temporal Data Model//Proc. of First East-European Symposium on Advances in Databases and Information Systems-ADBIS'97. St. -Petersburg, 1997: 371-378
- [14] Kjetil N, Marit L, Lene M. TeXOR: Temporal XML Database on an Object-Relational Database System//Proc. of Perspectives of System Informatic. Novosibirsk, Russia, 2003
- [15] 潘顺, 金远平, 欧阳晔. 结构化数据到 XML 数据的语义映射. 东南大学学报(自然科学版), 2002, 32(3): 331-334
- [16] MA Z M. Fuzzy Databases Modeling with XML. Kluwer Academic Publishing, 2005
- [17] Lee D, Chu W W. Constraints-preserving transformation from XML documents type definition to relational schema. Lecture Notes in Computer Science, 1920, 2002: 339-353
- [18] Ji Y J, Tang Y, Ye Xiaoping. Study of Storage and Retrieval of Temporal XML Document. Journal of Information and Computational Science, 2004, 1(3): 117-122
- [19] Ye X P, Tang Y, Huang Z M. Query XML Documents with Temporal Variables//Proc. of the 8th Intl Conf. on CSCWD. Xiamen, 2003: 536-540
- [20] 周竞涛, 张树生, 孙宏伟, 等. 关系模式到 XML Schema 的约束保留映射. 西北工业大学学报, 2003, 21(3): 372-376