

# 有时态本体模型及其管理原型系统

田晓云, 方丽英, 闫健卓, 王 普

(北京工业大学电子信息与控制工程学院, 北京 100124)

**摘要:** 为使本体能够方便地表达、处理与时间相关的信息, 将有效时间概念引入本体模型, 介绍有时态本体模型定义及相关定理, 在此基础上, 实现有时态本体管理原型系统, 并将该原型系统应用于存在时态冲突的信息集成系统中进行实验。实验结果证明了将本体用有效时间进行扩展, 从而形成有时态本体模型, 并对其存储管理的必要性, 同时也显示了该原型系统对时态查询的支持能力。

**关键词:** 有时态本体; 有效时间; 原型系统

## Temporal Ontology Model and Its Management Prototype System

TIAN Xiao-yun, FANG Li-ying, YAN Jian-zhuo, WANG Pu

(College of Electronics Information and Control Engineering, Beijing University of Technology, Beijing 100124)

**【Abstract】** In order to express and deal with the time-related information using ontology, this paper introduces valid time into ontology and presents a kind of Temporal Ontology Model(TOM). It provides the definition and related theorems of TOM. On this basis, it implements a prototype of Temporal Ontology Management System(TOMS) and applies it to information integration systems which exists temporal conflicts. Experimental results show the necessity of temporal extensions and management to ontologies. It also demonstrates that TOMS has ability to support temporal queries.

**【Key words】** temporal ontology; valid time; prototype system

### 1 概述

已有研究关注本体中的变化、Web 服务中时态信息的重要性, 但对本体中的时态信息, 尤其是有效时间信息的表达、存储管理和查询等问题还不够重视。Abiteboul 指出对于时间的建模是 Web 和半结构化数据查询语言中最为关键的一个问题。因此, 作为语义网<sup>[1]</sup>层次模型中的核心层, 如果本体模型能够扩展到支持表达随时间进展而变化的知识以及这些知识所包含的时态信息, 将能更好地反映本体的动态性特征。Gutierrez 等人首次把时态数据库中的有效时间概念引入 RDF, 提出时态 RDF 的概念。对 RDF 中的时态信息进行建模和查询, 并在文献[2]中给出完整的推导和证明。文献[3]提出 OWL 的时态扩展, 为 OWL 添加新标签, 用以描述概念的新旧版本及版本变化时的事物时间。此外, 针对本体时间信息的展示方面, 文献[4]针对本体沿时间轴的演变, 为 protégé 增加可视化插件, 用以展示本体中实例的时间信息。

本文根据 OWL 所描述内容的有效时间提出有时态本体模型, 以形成通用的管理模型及查询机制。

### 2 有时态本体模型

类比于时态数据库中有有效时间的概念, 本文定义本体元素的有效时间是指该本体元素所表示的知识在本体所描述的现实世界中发生并保持有效的那段时间。向本体及其元素引入有效时间的概念, 定义有时态本体模型如下。

#### 2.1 相关定义

**定义 1(有时态本体)** 有时态本体  $O_i$  为一个五元组:

$$O_i = (C[i], P[i], R[i], A[i], I[i]) \quad (1)$$

其中,  $O$  代表本体;  $C$  为概念集;  $P$  为属性集;  $R$  为关系集;  $A$  为公理集;  $I$  为实例集。在这些集合中, 每个元素都带有其

各自的有效时间标记  $t$ 。有时态本体  $O_i$  的有效时间  $t$  是所有这些本体元素的有效时间在时间轴上的并集。

由定义可知, 求取有效时间为  $[t_1, t_2]$  的本体片段  $O[t_1, t_2]$ , 就是求  $O_i$  中所有有效时间落于  $[t_1, t_2]$  时段区间之内的元素的子集, 即  $O[t_1, t_2]$  定义为

$$O[t_1, t_2] = O(\langle C, P, R, A, I \rangle : [t_1, t_2] : t_1 \leq t_2) \quad (2)$$

特别地, 当  $t_1 = t_2$  时, 则式(2)就转化为本体在  $t_1 = (t_2)$  时刻的快照, 即  $O_{i,t_2} = O_{i,t_1}$  为无时态的本体。

**定义 2(有时态本体元素)** 在有时态本体中, 各个本体元素记为

$$\{item_{ij} : [t_{ij1}, t_{ij2}] \mid i \in N, j \in M, t_{ij1} \leq t_{ij2}\} \quad (3)$$

其中,  $i$  表示该元素在时态本体中的全局唯一 ID, 且时态本体中记录了该本体元素的  $j$  个有效时间段。

**定义 3(求有时态本体元素的有效时间)** 为了求有时态本体元素的有效时间, 定义  $validTime()$  方法。该方法是个从有时态本体元素到该元素相应的有效时间的映射, 定义域是有时态本体元素, 值域是时间实体。即有:

$$validTime(item_{ij} : [t_{ij1}, t_{ij2}]) = [t_{ij1}, t_{ij2}] \quad (4)$$

在有时态本体定义的基础上, 提出有时态本体模型中 2 个用于时间运算的基本定理。

**基金项目:** 北京市组织部优秀人才培养基金资助项目(110105197712 102924)

**作者简介:** 田晓云(1983—), 女, 硕士研究生, 主研方向: 语义搜索, 信息集成; 方丽英, 博士; 闫健卓, 副教授; 王 普, 教授、博士生导师

**收稿日期:** 2009-11-03 **E-mail:** fangliying@bjut.edu.cn

## 2.2 相关定理

**定理 1** 有时态本体  $O_i$  的快照本体求取方法为

$$O(t_0) = \{item_{ij_{min}} \mid \text{for each } i, j_{min} \text{ 满足 } \min\{t_0 - \text{beginningOf}(\text{validTime}(item_{ij} : [t_{ij1}, t_{ij2}]))\}\} \quad (5)$$

**定理 2** 有时态本体  $O_i$  在特定时间段  $[t_1, t_2]$  内的有时态本体  $O[t_1, t_2]$  的求取方法为

$$O(t_1, t_2) = U_i\{U_j\{item_{ij} : ([t_{ij1}, t_{ij2}] \cap [t_1, t_2])\}\} \quad (6)$$

## 3 有时态本体管理原型系统设计与实现

为证明有时态本体模型的可行性和有效性, 本文开发了有时态本体管理原型系统(Temporal Ontology Management System, TOMS)。

### 3.1 系统结构

有时态本体管理系统的结构是一个五元组  $\Delta = \{O_s, \Omega, T, \Gamma, O_{TS}\}$ , 其中,  $O_s$  为本体集, 采用 OWL 作为表示语言, 存储于 OWL 文件中:  $O_s = \cup O_i$ ;  $\Omega$  为知识集, 即本体以三元组  $(p, s, o)$  为表示方法的知识集, 有:  $\Omega = \{U(p, s, o) \mid \forall (p, s, o)_{ij} \in O_i, O_i \in O_s\}$ ;  $T$  为时间实体, 本文采用基于区间的时间元素, 即通过描述时间段的起始点和中止点来描述时间区间, 有:  $T = \{[t_i, t_j] \mid t_i \leq t_j, ij \in N\}$ ; 当  $t_i = t_j$  时,  $T$  为时间点;  $\Gamma$  为有时态知识集, 有  $\Gamma \subseteq \Omega \times T$ , 其具体表示为五元组, 其中  $[t_{start}, t_{end}]$  为知识集的有效时间区间:  $\Gamma = \{U(p, s, o, t_{start}, t_{end})_{ij} \mid \forall (p, s, o)_{ij} \in O_i, O_i \in O_s\}$ 。  $O_{TS}$  为有时态本体集, 其知识库源于  $\Gamma$ , 采用 Time-Ontology 方式扩展 OWL 描述语言, 存储于 OWL 文件中。

### 3.2 系统设计

本文设计的 TOMS 具有以下功能: 实现有时态本体的解析和存储; 支持对于不同时间查询条件的有时态本体查询。

根据以上功能需求, TOMS 中设计有 4 个功能模块:

- (1) 有时态本体解析模块(TOPM): 从有时态本体集  $O_{TS}$  解析出知识集  $\Omega$ 。
- (2) 有时态本体存储模块(TOSM): 将有时态知识集  $\Gamma$  进行持久化, 存储至数据库。
- (3) 有时态本体构建模块(TOCM): 在查询过程中, 将满足查询条件的有时态知识集  $\Gamma$ , 构建为有时态本体集  $O_{TS}$ 。
- (4) 时态信息处理模块(TTM): 处理有效时间信息。在有时态本体存储过程中, 该模块从  $O_{TS}$  中提取它的有效时间实体  $T$ 。在查询过程中, 基于不同的时间查询条件, 与存储至数据库中的有效时间实体  $T$  进行匹配。

### 3.3 系统实现

在 TOMS 中, 输入/输出参数为有时态本体。以下为采用 Time Ontology 表示有效时间的有时态本体的 OWL 描述模板。其中, Statement 为陈述; StartTime 和 EndTime 分别为陈述有效时间的开始时间和结束时间。

```
<rdf:RDF xmlns:t="http://www.w3.org/2006/time#">
<t:Interval rdf:about=%Statement%>
<t:hasBeginning rdf:resource=%StartTime% />
<t:hasEnd rdf:resource=%EndTime% />
</t:Interval>
```

TOMS 采用 Java 语言开发(JDK6.0)。在 OWL 文件与三元组的转换部分, 调用惠普(HP)语义网实验室的 JenaAPI。TOMS 将有时态本体持久化至 Oracle 数据库, 时态信息的管

理采用 TimeDB 插件进行扩展。

## 4 有时态本体管理原型系统的应用实验

近年来, 本体被广泛用于信息集成系统中, 主要用于对数据源语义的说明描述。

信息集成系统需要集成的数据源大多是在不同的历史时期以不同的角度独立设计部署的, 这种差异必然会带来与时间相关的语义冲突。同时, 随着集成系统的深入使用, 新加入集成系统的数据源, 其包含的知识可能与已有信息的语义相冲突。设计将 TOMS 应用于信息集成系统中, 不但可以正确地表示知识的时间差异, 同时可解决知识随时间演变带来的语义冲突。TOMS 在信息集成系统中的应用原理, 如图 1 所示。

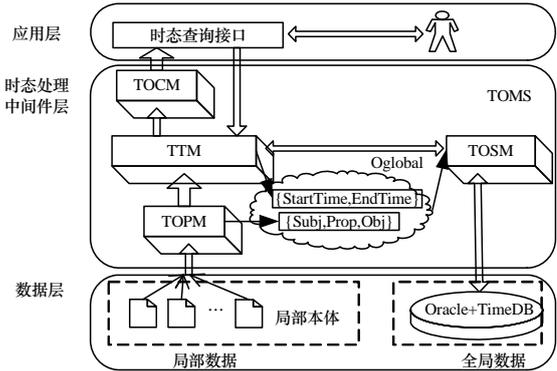


图 1 系统体系结构

局部本体是对局部数据源的概念描述, 将其输入至 TOMS。TOPM, TTM 模块对输入的局部本体进行解析, 提取知识及其有效时间, 并根据有效时间处理语义冲突, 对知识进行概念筛选、重构, 得到知识内容最为全面且运用有时态本体模型表达的全局有时态本体 Oglobal, 最后 Oglobal 经 TOSM 模块持久化至时态数据库。当人或应用系统发出查询请求时, 首先由 TTM 模块对查询请求的时间条件进行分析, 根据相应定理对持久化至时态数据库中的 Oglobal 进行时态匹配, 将满足条件的概念、关系、实例等集合返回, 由 TOCM 重构成有时态本体片断或无时态的本体快照, 以 OWL 格式输出。

## 5 实验结果与分析

实验以在不同时间段建立的 3 个数据源  $D_1, D_2, D_3$  (分别建立于 1994 年底、2000 年初和 2005 年初) 为局部数据源, 其内容包括北京工业大学的院系设置信息。针对这 3 个数据源分别建立 3 个局部本体  $O_1$  (图 2)、 $O_2$  (图 3)、 $O_3$  (图 4)。

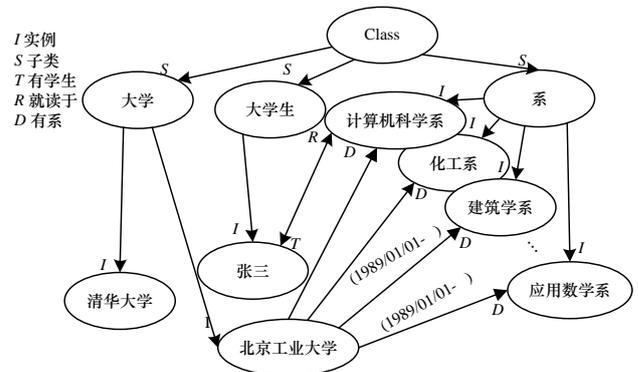


图 2 局部本体  $O_1$

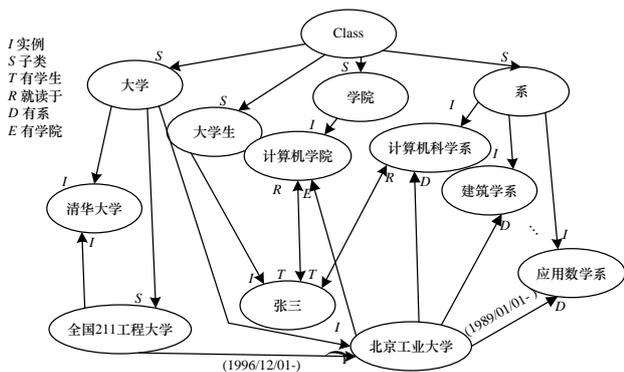


图3 局部本体  $O_2$

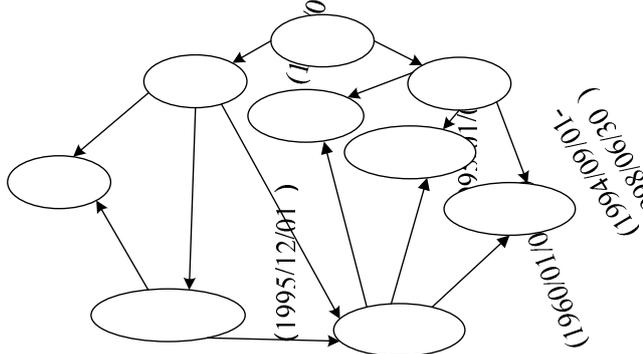


图4 局部本体  $O_3$

利用 Time Ontology 对其中的有效时间信息进行预处理，形成有时态本体管理原型系统的输入数据，包含有效时间信息的局部本体文件。

### 5.1 时态存储管理

有时态本体管理原型系统将  $O_1, O_2, O_3$  相对应的 3 个 OWL 文件解析、处理，形成五元组形式的全局时态本体 Oglobal(如图 5)，并持久化保存到时态数据库中。

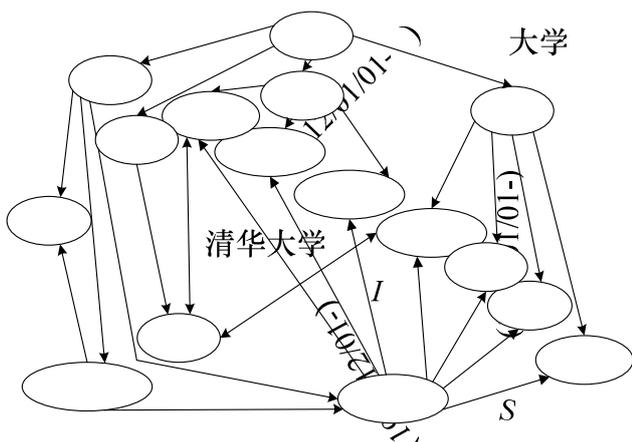


图5 全局本体 Oglobal

从图 5 可以看到，Oglobal 中包含 3 个局部本体中的全部知识。同时通过对时态信息的表示和处理，解决了集成过程中的语义冲突。如计算机科学与计算机学院在 Oglobal 中并存，但由于其有效时间不同，消除了知识的语义冲突。

### 5.2 时态查询

**实验 1** 设计求 Oglobal 在给定时间点的无时态本体快照的实验。向 TOMS 输入时间点 2004-02-05 的本体快照查询请求，TOMS 输出了 OWL 描述的本体快照。为了清晰展示，

给出三元组形式的结果片段：

北京工业大学 type 全国 211 工程大学  
北京工业大学 有系 建筑学系…  
北京工业大学 有学院 计算机学院

查询结果显示，北京工业大学在给定时间点属于 211 工程院校；其院系结构包括计算机学院，无计算机科学系。作为实验结果的本体切片所描述的知识与给定时间点的客观现实相符合。本实验测试了 TOMS 求取有时态本体快照的功能，同时也验证了文中定理 1 的正确性。

**实验 2** 设计实验求 Oglobal 在给定时间段的有时态本体片段。向 TOMS 输入时间段 1988-01-01~2004-02-05 的本体片段查询请求，TOMS 输出 OWL 表述的有时态本体，以下是五元组形式结果片段：

1989/1/1 1989/12/31 北京工业大学 有系 化工系  
1989/1/1 1989/12/31 北京工业大学 有系 计算机科学系…  
1998 2004/2/6 北京工业大学 有学院 计算机学院  
2004 2004/2/6 北京工业大学 有学院 应用数理学院

在 1988-01-01~2004-02-05 期间院系设置发生了一系列变化，从实验结果可看到有效时间段处于此范围内的所有本体元素及其时态信息。实验测试了 TOMS 求有时态本体在指定时间段内的有时态本体片段的功能，同时也检验文中定理 2 的正确性。

**实验 3** 为了突出 TOMS 对带有有效时间的查询的支持，设计查询给定 Subject 沿时间轴变化的实验。针对 Subject=“张三”的查询请求，输出结果为：

[1994/9-1998/6/30] http://university.owl#张三  
http://www.w3.org/1999/02/22-rdf-syntax-ns#type  
http://university.owl#大学生  
[1994/9-1997/12/31] http://university.owl#张三  
http://university.owl#就读于  
http://university.owl#计算机科学系  
[1998-1998/6/30] http://university.owl#张三  
http://university.owl#就读于  
http://university.owl#计算机学院

实验结果显示张三曾就读于计算机科学系和计算机学院。如果不记录有效时间，通常的全局查询将会得到“张三就读于计算机科学系”或“张三就读于计算机学院”的片面信息。即使查询结果能够并列罗列前述 2 个结果，由于缺乏有效时间的标示，用户也必然会对不一致的输出表示困惑。可见，由于保存了全局有时态本体，因此 TOMS 支持查询本体元素随时间的变化过程。

**实验结论**  
在全局存储有时态本体表示的全局模式，其所能支持的查询与传统信息集成系统的全局查询相比，能更有效且直观地处理时态相关信息。上述实验表明了在本体中标示有效时间的必要性。同时，将 TOMS 应用于信息集成系统中可解决对于存在时态差异的知识表示、存储和查询问题。

### 6 结束语

为了使本体能够清晰地表达和处理时态信息，本文将有效时间引入本体，提出了有时态本体模型，并给出针对此模型开发了有时态本体管理原型系统 TOMS。进一步引入事物时间，实现本体的双时态表示及其管理是下一步工作重点。

(下转第 50 页)