

时态知识 数据模型研究及应用^{*}

汤 庸, 汤 娜, 毛承洁, 叶小平, 余 阳, 梁 路

(中山大学计算机科学系, 广东 广州 510275)

摘 要: 时间是信息的重要属性, 不仅数据有时态性, 而且知识也有时态性。时态信息在信息系统中扮演着日益重要的角色, 在某些系统中还起着关键性作用。将时态应用分为完全时态应用、嵌入式时态应用和混合型时态应用3种模式, 提出一种时态信息模型, 重点讨论时态数据模型和时态知识表达模式, 介绍典型时态信息系统“工资智能决策支持系统”中的知识/数据处理模型。

关键词: 时态数据模型; 时态知识表达模式; 时态应用模式; 工资系统

中图分类号: TP311 **文献标识码:** A **文章编号:** 0529-6579 (2004) 06-0062-05

时间是自然界无所不在的客观属性, 时态信息技术实际上一直伴随着数据库技术的发展而产生和发展, 它经历了3个主要时期^[1,2]: 1982年以前的开创期, 加州大学洛杉矶分校 Ben Zvi 和纽约大学的 Clifford 在 1982 年的博士论文是时态数据库技术开创期的2个标志性成果, 最重要的贡献是提出以时间区间作字段值, 引入了有效时间(valid time)和用事务时间(transaction time)双时态(bi temporal)概念^[3,4]。1983—1993年的发展繁荣期, 这个时期时态数据库的理论模型已基本形成^[5]。1994年以后进入了时态信息技术的第3个时期, 如何将时态数据模型“标准化”和“产品化”是这个时期重要特征, 该时期另一个重要特征是“时态信息的应用”。随着数据库与信息技术的深入和发展, 信息系统面临许多新的应用和新的需求, 对时态信息处理的需求越来越迫切。近年来, 时态信息技术研究与应用越来越受到人们的关注, 但是成熟的时态信息处理技术和软件产品还很少见, 特别是时态知识模型还未见报道。本文给出一种时态应用开发模式的分类方法, 提出时态信息模型(TIM), 介绍一种典型时态信息系统, 最后指出进一步研究工作。

1 时态应用模式

按照应用系统是否处理时态属性, 我们将应用系统分为非时态应用(传统的应用系统)和时态应用系统。将不支持时态属性信息的处理的传统应用

称为非时态应用系统; 提供时态信息处理功能的应用系统称为时态应用系统(temporal application); 又根据系统处理时态属性的方式和能力, 将时态应用系统分为3类: 完全时态应用系统、嵌入式时态应用系统和混合型时态应用系统(见图1)。

1.1 混合式时态应用模式

混合式型时态应用模式(mixed temporal application mode)是指将时态数据处理技术与应用信息处理技术结合, 实现应用系统中对时态属性的支持, 其底层数据库系统也是传统的 DBMS, 见图1(b)。目前大多数时态应用系统都属于这种类型, 本文介绍的实例“工资智能决策支持系统”也属于这类模式。

1.2 嵌入式时态应用模式

嵌入式时态应用模式(embedding temporal application mode)是通过一些时态处理开发工具(时态中间件)来实现应用系统中的时态信息处理功能, 底层数据库系统采用传统的 DBMS, 见图1(c)。

这类系统是开发时态应用的方向。开发这类模式的应用需要时态处理工具软件支持, 例如 Time Consult 公司的 Time DB、我们开发的 TIM 等都是这样一些时态处理中间件。

在逻辑上, 可以将嵌入式时态应用软件分为3个层次, 如图1(c)所示。

(1) 时态应用层: 此层次是和用户联系的窗口, 它提供了应用时态预处理工具来定义时态数据视图,

* 收稿日期: 2004-09-16

基金项目: 国家自然科学基金资助项目(60373081); 广东省自然基金重点资助项目(04105503); 广东省十五重大专项基金资助项目(A010203); 广州市重点科技资助项目(2002Z2-D3041); 广东省跨世纪优秀人才基金资助项目(Q02095)

作者简介: 汤 庸 (1964年生), 男, 教授, 博士生导师, E-mail: jssy@zsu.edu.cn

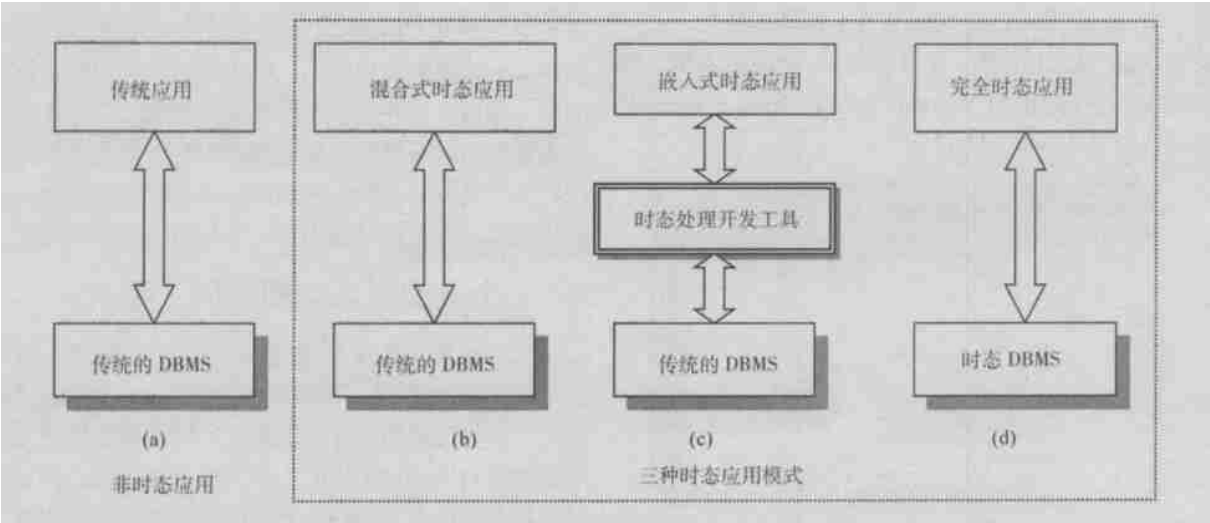


图 1 时态应用模式分类

Fig 1 Three temporal application modes

编辑时态知识库；调用时态中间件服务，完成时态信息处理。

(2) 时态中间层：提供基于 Web 跨数据库平台的时态信息处理服务，包括时态查询、时态推理等。

(3) 数据库层：实现在传统的数据库平台上时态信息的表示和存储。

1.3 完全时态应用模式

完全时态应用模式也称为纯粹时态应用模式 (Pure Temporal Application mode)，是指基于时态数据库系统开发的应用系统，它需要完全的时态数据库支持，其底层数据库本身是时态数据库管理系统 TDBMS，见图 1 (d)。目前还没有时态数据库管理系统 (TDBMS) 商业产品，主流数据库管理系统都没有支持时态数据管理的功能，完全时态应用模式的应用系统目前还未见报道。

2 时态信息模型

时态信息处理技术或工具是开发时态应用系统的核心技术，时态中间件研究已成为时态开发工具重要发展方向。我们提出一种时态信息模型 (temporal information model)，简称 TIM^[1]。时态信息模型 TIM 主要包括 2 个部分：时态信息处理软件构件 (temporal information processing software component) 和时态知识/数据模型 (temporal knowledge/data model)。这里，时态信息 (temporal information) 包括时态数据 (temporal data) 和时态知识 (temporal knowledge)。

2.1 时态数据模型

时态数据模型 TDM (temporal data model) 是建立时态信息系统的基础。在 20 世纪 80—90 年代形成了许多时态数据模型，从早期的基于关系代数，到后来

的 calculus based, datalog based, OO 等。经过近年来的发展与实践，学术界的观点和认识逐步趋向“统一”，现在基本上采用扩充的 SQL 模型，即 Snodgrass 在 1996 年提交的 SQL/Temporal 标准^[5,6]。

目前大部分时态数据查询语言是扩展了传统的数据库查询语言如 SQL 或 Quel，主要方法是在 TDB 中增加一些运算，如 AFTER、BEFORE 和 OVERLAP 等；扩展一些操作，如时态选择、时态投影、时态连接等。这些扩展后的时态数据查询功能有限，效率较低。由于种种原因，数据库厂商难以下决心选用 TDB 技术用于产品。

一般地，在时态应用中包括非时态信息与时态信息两类。在模式 2 (图 1(c)) 的基于传统数据库管理系统开发时态应用时，由于传统的数据库没有支持时态属性的机制，所以“无缝”连接非时态信息与时态信息是关近技术之一。我们提出采用时态索引 (temporal index) 实现逻辑上的“统一”时态视图 (temporal view)，实现非时态信息与时态信息的“无缝”连接。具体方法为将字段分为传统的数据字段 (非时态数据字段) 和时态数据字段 (时态数据字段为时态数据类型) 2 组，非时态字段采用传统的数据表 (非时态数据字段) 实现，时态数据字段扩充一个新的数据表 (包含时态数据字段)，2 张表的连接采用时态索引来实现，即在数据库的内采用 2 张表，然后基于这 2 张表定义一个时态视图，用户所有的操作都基于这个时态视图来进行，在时态处理中间件中将其翻译转化为对实际物理表的操作。采用这种索引的方法不会改变传统数据库的定义与操作，保持了对传统数据库操作的基本特性，易于得到广泛接受。目前，我们已初步建立时态处理中间件，基于 JAVA 实

现了 Temporal SQL 语句到标准 SQL 语言的转化, 底层采用 ORACLE8. 16. 图 2 是 TDM 的一个演示示例, 图 2 是将时态查询语句 Validtime select *from p (period) 转化后生成的标准 SQL 语句^[7]。



图 2 TDM 演示示例
Fig. 2 A demo of TDM

2. 2 时态知识 /模型

目前时态数据库研究大多局限在数据库的时态属性, 而忽视了其他信息的时态属性, 例如知识库的时态特征。时态数据技术还停留在“数据”处理上, 符号演算和推理能力强, 但信息处理能力弱。其原因是时态逻辑和推理方面的研究与时态数据库和时态信息处理研究还没有接轨。

时态知识模式 (temporal knowledge mode) 的基本方法也是在传统的知识表达模式 (产生式规则、框架、语义网、面向对象等) 的基础上引入双时态概念, 将有效时间引入知识单元 (例如一个知识对象、一条规则), 将事务时间引入于知识推理中, 解决时态冲突, 实现逆向推理等。下面是一种“规则+过程式”时态规则的形式化描述:

Rule Name;
Condition; T-tempression
ACT: T-procedure1
T-Procedure2
.....
T-Procedure2
Valid-Time; Time-value
Other Attributes

时态知识库应建立在时态数据库的基础上, 因为如果数据没有时态性, 时态知识也就没有应用的价值。所以时态知识与时态数据之间应有“统一”的桥梁。在上述形式化描述中, T-tempression 是时态表达式, 由时态函数 (T-Function)、时态算子、时态

变量与常量组成。推理机制通过 T-function 和 T-procedure 读取时态数据库, 实现时态知识和时态数据的统一处理。在开始一个推理过程之前, 首先要根据时态数据的时态属性值来选取相匹配的知识规则, 匹配的标准是将时态数据和知识的两者时态属性做时态逻辑运算, 若运算结果为真, 则与知识规则匹配^[8,9]。

3 一个典型时态应用实例

工资智能决策支持系统 (SIDSS) 也称为工资时态知识数据库系统, 是我们开发的一个实际应用系统^[10, 11]。与传统的财务工资系统不同, 该系统使基于有关政策自动确定工资的人事工资智能系统, 其目标是将各单位的人事部门从繁琐的工资调整工作中解放出来, 不再陷于成堆的工资政策中, 同时也可以避免许多人为造成的错误, 使工资管理规范。系统已成功推广应用到近万多家用户, 在全国率先实现公务员工资财政统一发放。

SIDSS 是一个典型的时态应用系统, 其时态性主要表现在: 不但人员信息、工资数据具有时态性, 工资政策、工资标准等还具有鲜明的时态知识特征。一个人员的工资需要根据他的学历、职务及其获得时间等信息 (时态性数据), 依据相应时期的工资政策 (时态性知识) 来确定。图 3 是工资系统时态知识/数据框架, 阴影框“工资日常管理”、“时态数据库管理”、“时态知识库管理”和“时态工资变更”是系统的功能模块, 双线框是时态中间处理程序。下面仅以几个实例简要讨论 SIDSS 时态信息现象, 详细技术细节不在本文介绍。

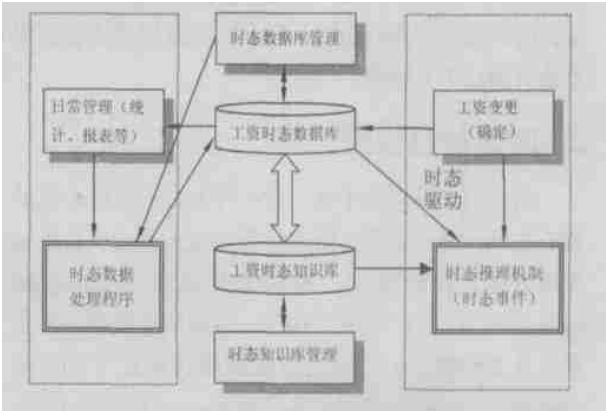


图 3 时态工资系统中时态知识/数据模型
Fig. 3 The temporal knowledge/data model of temporal salary system

3. 1 时态工资数据

工资系统包括 3 类信息: 第 1 类是不影响工资的信息, 如人员编号、姓名、身份证号等; 第 2 类是影响工资但不受时间影响的信息, 如参加工作时间等,

第 3 类是影响工资也受时间影响的信息，如学历、职务晋升、工资标准等。另外，还有在工资确定过程中产生的人员工资历史记录，它们同样具有时态性。工资管理决策支持系统的一大重要特点是根据用户输入的人员信息重现该人员自参加工作以来的工资变动历程，以确保工资确定的正确性。由此可知，SIDSS 既要记录有效时间，又要记录事务时间。工资管理决策支持系统的一大重要特点是根据用户输入的人员信息重现该人员自参加工作以来的工资变动历程，以确保工资确定的正确性。影响工资变动的信息主要包括：参加时间、学历、职务级别、任职时间、考核结果、奖励及处分等。工资时态是三维数据实体，用传统的关系数据库（二维表）和 1NF 模式难以表达，我们扩充 TER（时态 ER）模型建立时态工资数据非 1NF 元组^[10,11]。表 1 是一个时态数据示例。

3.2 时态工资标准的结构化知识描述

所谓结构化的时态知识，是指结构比较明确的知识，这种知识有时态特征。例如：讲师的职务工资标准为：“从 1993 年 10 月 1 开始，讲师一档至五档的职务工资分别为：210、228、246、264、312 元；从 1997 年 7 月 1 日开始，讲师一档至五档的职务工资分别为：226、244、262、304、328。…”

这是一条典型的结构化工资政策，它有以下几个特点：① 数据结构明确；② 知识中包含着数据；③ 知识的时态特征主要表现为有效时间。

```
该类知识对象可以描述为：
public class TK 讲师工资标准 {
    private String sTK-Id;           //对象标识
    public Float sTK-档次 1;
    .....
    public Float sTK-档次 5;
    private IntervalBasedTime ValidTime //知识的有效时间
    Public Boolean TConstraint {      //时态约束
TConstraint=
Before (Prior. ValidTime, this. ValidTime) & Before (this.
ValidTime,
        Next. ValidTime)
    }
```

表 1 时态工资数据库示例

Tab. 1 An example of temporam salary database

人员编号	专业技术职务	部门	...
0101010001	[1993 7, 1996 10] 助教	[1993 7, 1995 5] 数学系	...
	[1996 11, 2000 10] 讲师	[1995 6, 1997 8] 计算机系	
	[2000 11, NOW] 副教授	[1997 9, 2000 2] 数学系	
		[2000 2, NOW] 网络中心	
0101010002	[1997 7, 2000 10] 助教	[1997 7, NOW] 机电系	...
	[2000 11, NOW] 讲师		
...	...	...	...

}
在上述定义中，把与时间相关的属性独立出来，这样考虑是为了对时态属性进行时态逻辑约束。记录的“执行时间”应该存在时序约束关系：本条工资标准的执行时间，必须介于前一条工资标准和下一条工资标准的执行时间之间。即：

```
Before ( Prior. ValidTime, this. ValidTime )
&Before ( this. ValidTime, Next. ValidTime)
```

3.3 工资政策的非结构化时态知识描述

工资政策是一种典型的时态知识，其有效性与时间紧密相关，人们最常见的是：“某某政策从 XX 年 XX 月 XX 日开始执行”。作为一种领域知识，它本身往往比较复杂，非结构化的特点很突出，能较好地描述这种知识，对表达其它领域的知识是一种启发和借鉴。

例如，“两年优晋升”的政策：1999 年以前，如果人员连续 2 年考核优秀，或其中某年考核优秀为“一年优”，工资档次允许晋升一档，考核结果有效年份从下年算起，但连续 2 次晋升之间必须相隔 2 年以上。1999 年以后的政策改为：连续两次晋升之间必须相隔 4 年以上。用面向对象的时态知识表达方法来描述这条知识，首先要仔细分析它隐含的判断条件：

我们把知识中隐含的条件分成 2 类，前一类只能用过程性函数来描述，后一类可以用时态算子描述，这样处理的目的在于最大限度地实现知识和代码的分离，如果关于时间的约束发生变化，则只需修改包含时态算子的表达式即可，而不必对应用系统的做太多改动。定义如下：

- (1) Boolean ValidNonimation (ID)，判断是否满足两年考核结果为“优”或“一年优”；
- (2) MGet -上一次晋升年份：用于返回上一次晋升的年份 Prior；
- (3) 晋升的时态约束表达式 TC Exp:
 $(GAP (Prior, Now) \geq 2 \text{ Year AND Before } (Now, 1999)) \text{ OR } (GAP (Prior, Now) \geq 4 \text{ Year AND After } (Now, 1999))$

TC_Exp 返回布尔结果给 Flag。
整条政策描述的结构见表 2。

表 2 时态工资政策知识对象描述示例
Tab. 2 The description of temporal salary policy knowledge object

对象结构	结构内容
STK-Id	TK-晋升
STK-Description	工资档次晋升
ValidTime	[1993—10—1 , + ∞]
Priority	1
TK-Trigger	ValidNomination (ID)
TK-Rule	TKRule-工资晋升
Method	MGet-上一次晋升年份; M-确定晋升一档

4 结 语

时态信息处理已成为许多新一代数据库与信息系统的核心技术，特别是电子政务、电子商务、数据仓库、数据挖掘、决策支持系统等信息系统中扮演着日益重要的角色，时态信息技术越来越受到人们的关注。但是，时态信息理论和技术处于发展阶段，成熟的软件产品还很少见。本文提出的时态知识/数据模型及其框架原型，已在典型时态信息领域得到成功应用，但是还没有形成平台级“时态中间件”产品。我们进一步研究的内容主要包括：双时态演算体系、时态数据库技术（时态索引、时态主键等）、时态知识表达与推理的库形式化描述（基于时态数据库的时态逻辑工具），另外还有时态信息的一些应用技术，如基于时态数据库的时态 workflow、时态协同工作机制、时态数据挖掘等相关理论和应用技术，实现一系列“时态中间件”产品。

参考文献:

[1] 汤庸. 时态数据库导论[M]. 北京: 北京大学出版社, 2004.

[2] 汤庸, 叶小平, 汤娜. 时态信息处理技术研究综述[J]. 中山大学学报(自然科学版), 2003, 42(4): 5—9.

[3] JENSEN C S. Introduction to temporal database research outline [EB/OL]. <http://infolab.usc.edu/csci599/Fall2001/presentations/session16.pdf>, Fall, 2001.

[4] SKYT J, JENSEN C S. A foundation for vacuuming temporal databases [J]. Data and Knowledge Engineering, 2003, 44(1): 1—29.

[5] TANSEL A, CLIFFORD J, GADIA J S, et al. Temporal databases: theory, design, and implementation. Database Systems and Applications Series [M]. Benjamin/Cummings, 1993.

[6] SNODGRASS R T. The TSQL2 temporal query language [M]. Kluwer Academic Publishers, 1995.

[7] TANG Y, LIANG L. Bitemporal extensions to non-temporal RDBMS in distributed environment // The 8th International Conference on Computer Supported Cooperative Work in Design, VolII [C]. Xiamen, 2004: 370—373.

[8] 汤庸, 汤娜, 叶小平, 等. 时态知识和时态数据的统一模型研究 [J]. 软件学报, 2003 (增刊): 74—78.

[9] LIU D L, TANG Y. An inference model of temporal logic in an intelligent decision support system of salary [C]. CSCWD 02, Rio de Janeiro, Brazil, 2002: 351—355.

[10] WEI W, TANG Y. The design of temporal knowledge base in SIDSS [C]. Proceedings of 4th International Conference on CAID&CD, 2001: 831—836.

[11] TANG Y, TANG N, MAO C J. Knowledge representation of intelligent decision support system of salary // Proceedings of The 7th international conference on IE&EM [C]. 2000: 519—523.

A Temporal Knowledge/Data Model and Its Application

TANG Yong, TANG Na, MAO Chengjie, YE Xiaoping, YU Yang, LIANG Lu
(Department of Computer Science, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: Time is an important attribute of information, which exists not only in database but also in knowledge base. Temporal information plays more and more important role in modern information systems, even decisively in some systems. Temporal applications are divided into three modes, the mix temporal mode, the embedding temporal mode and the pure-temporal mode. And a temporal information model is proposed, in which most attention is paid to the temporal data model and temporal knowledge representation mode. Furthermore, the temporal data/knowledge processing model of a typical temporal information system “Salary Intelligent Decision Support System” is introduced.

Key words: temporal data model; temporal knowledge representation mode; temporal application mode; salary system