Managing Schema Evolution using a Temporal Object Model

Iqbal A. Goralwalla, Duane Szafron, M. Tamer Özsu
Laboratory for Database Systems Research
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada T6G 2H1
{iqbal,duane,ozsu}@cs.ualberta.ca

and

Randal J. Peters
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba, Canada R3T 2N2
randal@cs.umanitoba.ca

Abstract

The issues of schema evolution and temporal object models are generally considered to be orthogonal and are handled independently. This is unrealistic because to properly model applications that need incremental design and experimentation (such as CAD, software design process), the evolutionary histories of the schema objects should be traceable. In this paper we propose a method for managing schema changes by exploiting the functionality of a temporal object model. The result is a uniform treatment of schema evolution and temporal support for many object database management systems applications that require both.

1 Introduction

In this paper we address the issue of schema evolution and temporal object models. These two issues are generally considered to be orthogonal and are handled independently. However, many object database management systems (ODBMS) applications require both. For example:

- The results reported in [Sjø93] illustrate the extent to which schema changes occur in real-world database applications such as health care management systems. Such systems also require a means to represent, store, and retrieve the temporal information in clinical data [KFT91, DM94, CPP95].
- The engineering and design oriented application domains (e.g., CAD, software design process) require incremental design and experimentation [KBCG90, GTC+90]. This usually leads to

To appear in the *Sixteenth International Conference on Conceptual Modeling (ER'97), November 1997.*

frequent changes to the schema over time which need to be retained as historical records of the design process.

We propose a method for managing schema changes by exploiting the functionality of a temporal object model. The provision of time in an object model establishes a platform from which temporality can be used to investigate advanced database features such as schema evolution. Given that the applications supported by ODBMSs need support for incremental development and experimentation with changing and evolving schema, a temporal domain is a natural means for managing changes in schema and ensuring consistency of the system. The result is a uniform treatment of schema evolution and temporal support for many ODBMS applications that require both.

Schema evolution using time is the process of allowing changes to schema without loss of information. Typical schema changes include adding and dropping behaviors (properties) defined on a type, adding and dropping subtype relationships between types, to name a few. Using time to maintain and manage schema changes gives substantial flexibility in the software design process. It enables the designers to retrieve the interface of a type that existed at any time in the design phase, reconstruct the super(sub)-lattice of a type as it was at a certain time (and subsequently the type lattice of the object database at that time), and trace the implementations of a certain behavior in a particular type over time.

A typical schema change can affect many aspects of a system. There are two fundamental problems to consider:

1. **Semantics of Change**. The effects of the schema change on the overall way in which the system organizes information (i.e., the effects on the schema). The traditional approach to solving this problem is to define a set of invariants that must be preserved over schema modifications.

2. **Change Propagation**. The effects of the schema change on the consistency of the underlying objects (i.e., the propagation of the schema changes to the existing object instances). The traditional approach of solving this is to *coerce* objects to coincide with the new definition of the schema.

In this paper we primarily consider the consistent handling of the problem of semantics of change using a temporal ODBMS. We describe the necessary modifications that could occur on the schema, and show how the implications of the modifications are managed. Our work is conducted within the context of the TIGUKAT[1] *temporal* ODBMS [ÖPS+95, GLÖS95, GÖS97] that is being developed at the University of Alberta. However, the results reported here extend to any ODBMS that uses time to model evolution histories of objects.

The remainder of the paper is organized as follows. In Section 2, we examine some of the previous work on schema evolution. In Section 3, we give a brief overview of the TIGUKAT temporal object model with an emphasis on how histories of objects are maintained. In Section 4, we describe the schema changes that can occur in TIGUKAT, and how they are managed using a temporal object model. In Section 5, we give examples of queries that allow software designers to retrieve schema objects at any time in their evolution histories. Concluding remarks and results of the paper are summarized in Section 6.

2 Related Work

The issue of schema evolution has been an area of active research in the context of ODBMSs [BKKK87, KC88, PS87, NR89]. In many of the previous work, the usual approach is to define a set of invariants that must be preserved over schema modifications in order to ensure consistency of the system. The Orion [BKKK87, KC88] model is the first system to introduce the invariants and rules approach as a more structured way of describing schema evolution in ODBMSs. Orion defines a complete set of invariants and a set of accompanying rules for maintaining the invariants over schema changes. The work of Smith and Smith [SS77] on aggregation and generalization sets the stage for defining invariants when subtypes and supertypes are involved. Changes to schema in previous works are *corrective* in that once the schema definitions are changed, the old definitions of the schema are no longer traceable. In TIGUKAT, a set of invariants similar to those given in [BKKK87] is defined. However, changes to the schema are not corrective. The provision of time in TIGUKAT establishes a natural foundation for keeping track of the changes to the schema. This allows applications, such as CAD, to trace their design over time and make revisions, if necessary.

There have been many temporal object model proposals (for example, [RS91, SC91, WD92, KS92, CITB92, BFG96]). In handling temporal information, these models have focussed on managing the evolution of real-world entities. The implicit assumption in these models is that the schema of the object database is static and remains unchanged during the lifespan of the object

[1]TIGUKAT (tee-goo-kat) is a term in the language of Canadian Inuit people meaning "objects." The Canadian Inuits, commonly known as Eskimos, are native to Canada with an ancestry originating in the Arctic regions of the country.

database. More specifically, the evolution of schema objects (i.e., types, behaviors, etc) is considered to be orthogonal to the temporal model. However, given the kinds of applications that an ODBMS is expected to support, we have exploited the underlying temporal domain in the TIGUKAT temporal model as a means to support schema evolution.

In the context of relational temporal models, Ariav [Ari91] examines the implications of allowing data structures to evolve over time in a temporal data model, identifies the problems involved, and establishes a platform for their discussion. McKenzie and Snodgrass [MS90] develop an algebraic language to handle schema evolution. The language includes functions that help track the schema that existed at a particular time. Schema definitions can be added, modified, or deleted. Apart from the addition and removal of attributes, the nature of the modifications to the schema and their implications are not demonstrated. Roddick [Rod91] investigates the incorporation of temporal support within the meta-database to accommodate schema evolution. In [Rod92], SQL/SE, an SQL extension that is capable of handling schema evolution in relational database systems is proposed using the ideas presented in [Rod91]. The approach used in the TIGUKAT temporal object model is similar in the sense that temporal support of real-world objects is extended in a uniform manner to schema objects, and then used to support schema evolution. Some of the ideas in [Rod91, Rod92, Rod95] have been carried forward in the design of the TSQL2 temporal query language [Sno95].

Skarra and Zdonik [SZ86, SZ87] define a framework within the Encore object model for versioning types as a support mechanism for changing type definitions. A type is organized as a set of individual versions. This is known as the *version set* of the type. Every change to a type definition results in the generation of a new version of the type. Since a change to a type can also affect its subtypes, new versions of the subtypes may also be generated. This approach provides fine granularity control over schema changes, but may lead to inefficiencies due to the creation of a new version of the versioned part of an object every time a single attribute changes its value. In our approach, any changes in type definitions involve changing the history of certain behaviors to reflect the changes. For example, adding a new behavior to a type changes the history of the type's interface to include the new behavior. The old interface of the type is still accessible at a time before the change was made. This alleviates the need of creating new versions of a type each time any change is made to a type.

3 The TIGUKAT Temporal Object Model

3.1 Fundamentals of TIGUKAT Object Model

The TIGUKAT object model [Pet94, ÖPS+95] is purely *behavioral* with a *uniform* object semantics. The model is *behavioral* in the sense that all access and manipulation of objects is based on the application of behaviors to objects. The model is *uniform* in that every component of information, including its semantics, is modeled as a *first-class object* with well-defined behavior. Other typical object modeling features supported by TIGUKAT include strong object identity, abstract types, strong typing, complex objects, full encapsulation, multiple inheritance, and parametric types.

The primitive objects of the model include: *atomic entities* (reals, integers, strings, etc.); *types* for defining common features of objects; *behaviors* for specifying the semantics of operations that may be performed on objects; *functions* for specifying implementations of behaviors over types; *classes* for automatic classification of objects based on type[2]; and *collections* for supporting general heterogeneous groupings of objects. Figure 1 shows a simple type lattice that will be used to illustrate the concepts introduced in the rest of the paper.

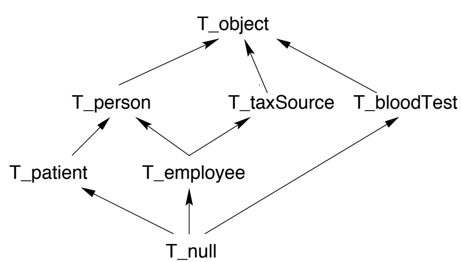

Figure 1: Simple type lattice.

In this paper, a reference prefixed by "`T_`" refers to a type, "**C_**" to a class, "*B_*" to a behavior, and "`T_X`< `T_Y` >" to the type `T_X` parameterized by the type `T_Y`. For example, `T_person` refers to a type, **C_person** to its class, *B_age* to one of its behaviors and `T_collection`< `T_person` > to the type of collections of persons. A reference such as joe, without a prefix, denotes some other application specific reference. The type `T_null` in TIGUKAT binds the type lattice from the bottom (i.e., most defined type), while the `T_object` type binds it from the top (i.e., least defined type). `T_null` is introduced to provide, among other things, error handling and null semantics for the model.

The access and manipulation of an object's state occurs exclusively through the application

[2] Types and their extents are separate constructs in TIGUKAT.

of behaviors. We clearly separate the definition of a behavior from its possible implementations (functions). The benefit of this approach is that common behaviors over different types can have a different implementation in each of the types. This provides direct support for behavior *overloading* and *late binding* of functions (implementations) to behaviors.

3.2 The Temporal Extensions

The philosophy behind adding temporality to the TIGUKAT object model is to accommodate multiple applications that have different type semantics requiring various notions of time [LGÖS97, GÖS97]. Consequently, the TIGUKAT temporal object model consists of an extensible set of primitive time types with a rich set of behaviors to model time. The only part of the temporal model that is relevant to this paper is the management of event histories. Therefore, we focus on history management and details of other aspects can be found in [GLÖS95, GLÖS96].

Our model represents the temporal histories of real-world objects whose type is `T_X` as objects of the `T_history⟨T_X⟩` type. For example, suppose a behavior *B_salary* is defined in the `T_employee` type. Now, to keep track of the changes in salary of employees, *B_salary* would return an object of type `T_history⟨T_real⟩` which would consist of the different salary objects of a particular employee and their associated time periods.

A temporal history consists of objects and their associated timestamps (time intervals or time instants). One way of modeling a temporal history would be to define a behavior that returns a collection of <**timestamp**, **object**> pairs. However, instead of structurally representing a temporal history in this manner, we use a behavioral approach by defining the notion of a *timestamped object*. A timestamped object knows its timestamp (time interval or time instant) and its associated value at (during) the timestamp. A temporal history is made up of such objects. The following behaviors are defined on the `T_history⟨T_X⟩` type:

$$
\begin{aligned}
B\_history: &\quad \texttt{T\_collection}\langle\texttt{T\_timeStampedObject}\langle\texttt{T\_X}\rangle\rangle \\
B\_timeline: &\quad \texttt{T\_timeline} \\
B\_insert: &\quad \texttt{T\_X}, \texttt{T\_timeStamp} \rightarrow \\
B\_remove: &\quad \texttt{T\_X}, \texttt{T\_timeStamp} \rightarrow \\
B\_validObjects: &\quad \texttt{T\_timeStamp} \rightarrow \texttt{T\_collection}\langle\texttt{T\_timeStampedObject}\langle\texttt{T\_X}\rangle\rangle \\
B\_validObject: &\quad \texttt{T\_timeStamp} \rightarrow \texttt{T\_timeStampedObject}\langle\texttt{T\_X}\rangle
\end{aligned}
$$

Behavior *B_history* returns the set (collection) of all timestamped objects that comprise the history. A history object also knows the timeline it is associated with and this timeline is returned by the behavior *B_timeline*. The timeline basically orders the timestamps of timestamped objects [GLÖS96]. The *B_insert* behavior accepts an object and a timestamp as input and creates a timestamped object that is inserted into the history. Behavior *B_remove* drops a given object from the history at a specified timestamp. The *B_validObjects* behavior allows the user to get the objects in the history that were valid at (during) a given timestamp. Behavior *B_validObject* is derived from *B_validObjects* to return the timestamped object that exists at a given time instant.

Each timestamped object is an instance of the `T_timeStampedObject⟨T_X⟩` type. This type represents objects and their corresponding timestamps. Behaviors *B_value* and *B_timeStamp* defined on `T_timeStampedObject` return the value and the timestamp (time interval or time instant) of a timestamped object, respectively.

Example 3.1 Suppose the type `T_patient` shown in Figure 1 represents different patients in a hospital. To represent a patient's blood test history over the course of a particular illness, the behavior *B_bloodTests* is defined on `T_patient` to return an object of type `T_history⟨T_bloodTest⟩`. Each blood test is represented by an object of the type `T_bloodTest`. Therefore, the history of the different blood tests undertaken by joe (an instance of `T_patient`) would then be retrieved using the behavior application joe.*B_bloodTests*. Let us call this history object bloodTestHistory. Now, suppose joe was suspected of having septicemia[3] and had diagnostic hematology and microbiology blood tests on 15 January 1995. As a result of a raised white cell count, joe was given a course of antibiotics while the results of the tests were pending. A repeat hematology test was ordered on 20 February 1995. To record these tests, three objects with type `T_bloodTest` were created and then entered into the object database using the following TIGUKAT behavior applications:

$$\begin{aligned}\mathsf{bloodTestHistory}.B\_insert(\mathsf{microbiology}, 15\ January\ 1995)\\ \mathsf{bloodTestHistory}.B\_insert(\mathsf{hematology1}, 15\ January\ 1995)\\ \mathsf{bloodTestHistory}.B\_insert(\mathsf{hematology2}, 20\ February\ 1995)\end{aligned}$$

If subsequently there is a need to determine which blood tests joe took in January 1995, this would be accomplished by the following behavior application:

$$\mathsf{bloodTestHistory}.B\_validObjects([1\ January\ 1995,\ 31\ January\ 1995])$$

This would return a collection of the two timestamped objects, {timeStampedMicrobiology, timeStampedHematology1}, representing the blood tests joe took in January 1995. The first timestamped

[3] An infection of the blood.

object would have **microbiology** as its value and the second would have **hematology1** as its value[4].

To assist in clarifying the contents and structure of a history object, we give a pictorial representation of **bloodTestHistory** in Figure 2.

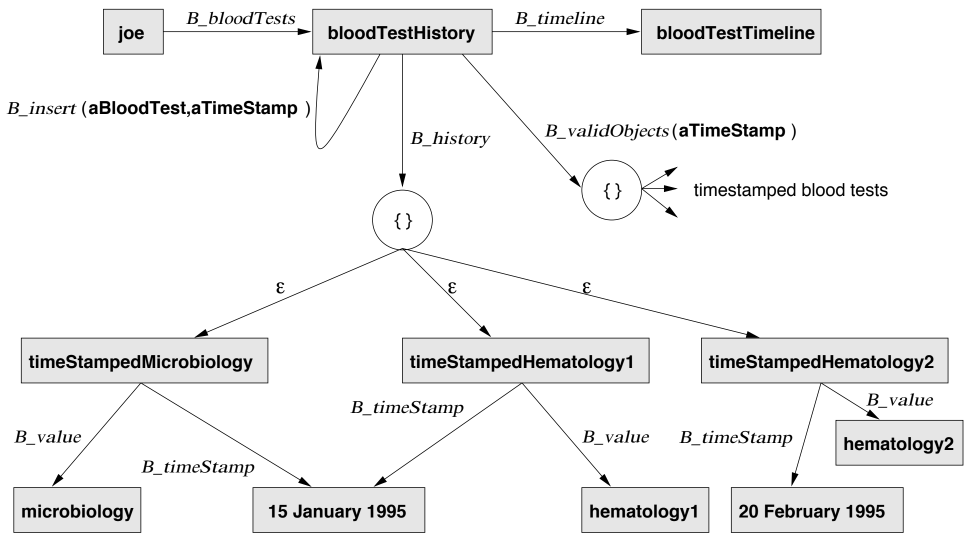

Figure 2: A pictorial representation of a patient's blood test history.

In the figure, the boxes shaded in grey are objects. Objects have an outgoing edge labeled by each applicable behavior that leads to the object resulting from the application of the behavior. For example, applying the behavior *B_timeline* to the object **bloodTestHistory** results in the object **bloodTestTimeline**. A circle labeled with the symbols { } represents a collection object and has outgoing edges labeled with "$\in$" to each member of the collection. For example, applying the *B_history* behavior to the object **bloodTestHistory** results in a collection object whose members are the timestamped objects **timeStampedMicrobiology**, **timeStampedHematology1**, and **timeStampedHematology2**. Finally, the *B_insert* behavior updates the blood test history (**bloodTestHistory**) when given an object of type `T_bloodTest` and a timestamp. Similarly, the *B_validObjects* behavior returns a collection of timestamped blood test objects when given a timestamp. □

[4]It should be noted that although we have two different timestamped objects containing the values microbiology and hematology1, they both contain the same timestamp. That is, although timeStampedMicrobiology.*B_value* = microbiology and timeStampedHematology1.*B_value* = hematology1, timeStampedMicrobiology.*B_timestamp* = timeStampedHematology1.*B_timestamp* = 15 *January* 1995.

4 Management of Schema Evolution by the Temporal Object Model

4.1 Schema Related Changes

There are different kinds of objects modeled by TIGUKAT, some of which are classified as schema objects. These objects fall into one of the following categories: *type, class, behavior, function,* and *collection*. There are three kinds of operations that can be performed on schema objects: *add, drop* and *modify*. Table 1 shows the combinations between the various schema object categories and the different kinds of operations that can be performed in TIGUKAT [Pet94, PÖ97]. The **bold** entries represent combinations that implicate schema changes while the *emphasized* entries denote non-schema changes.

	Operation		
Objects	Add (A)	Drop (D)	Modify (M)
Type (T)	**subtyping**	**type deletion**	**add behavior**(AB) **drop behavior**(DB) **add supertype link**(ASL) **drop supertype link**(DSL)
Class (C)	**class creation**	**class deletion**	*extent change*
Behavior (B)	*behavior definition*	**behavior deletion**	**change association**(CA)
Function (F)	*function definition*	**function deletion**	*implementation change*
Collection (L)	**collection creation**	**collection deletion**	*extent change*

Table 1: Classification of schema changes.

In the context of a temporal model, *adding* refers to creating the object and beginning its history, *dropping* refers to terminating the history of an object, and *modifying* refers to updating the history of the schema object. Since type-related changes form the basis of most other schema changes, we describe the modifications that affect the type schema objects. Type modification (depicted at the intersection of the M column and T row in Table 1) includes several kinds of type changes. They are separated into changes in the behaviors of a type (depicted as **MT-AB** and **MT-DB** in Table 1) and changes in the relationships between types (depicted as **MT-ASL** and **MT-DSL** in Table 1). Invariants for maintaining the semantics of schema modifications in TIGUKAT are described in [Pet94, PÖ97]. The invariants are used to gauge the consistency of a schema change in that the invariants must be satisfied both before and after a schema change is performed.

The meta-model of TIGUKAT is uniformly represented within the object model itself, providing reflective capabilities [PÖ93]. One result of this uniform approach is that types are objects and they have a type (called `T_type`) that defines their behaviors. `T_type` defines behaviors to access a type's interface (*B_interface*), its subtypes (*B_subtypes*), its supertypes (*B_supertypes*), plus many others that are not relevant for the scope of this paper. Since types are objects with well-defined behaviors, the approach of keeping track of the changes to a type is the same as that for keeping track of the changes to objects discussed in Section 3.2. This is one of the major advantages of the uniformity of the object model. The semantics of the changes to a type are discussed in the following sections.

4.2 Changing Behaviors of a Type

Every type has an *interface* which is a collection of behaviors that are applicable to the objects of that type. A type's interface can be dichotomized into two disjoint subsets:

1. the collection of *native* behaviors which are those behaviors defined by the type and are not defined on any of its supertypes;

2. the collection of *inherited* behaviors which are those behaviors defined natively by some supertype and inherited by the type.

There are three behaviors defined on `T_type` to return the various components of a type's interface: *B_native* returns the collection of native behaviors, *B_inherited* returns the inherited behaviors and *B_interface* returns the entire interface of the type.

Types can evolve in different ways. One aspect of a type that can change over time is the behaviors in its interface (i.e., adding or deleting behaviors). To keep track of this aspect of a type's evolution, we define histories of interface changes by extending the interface behaviors with time-varying properties. The definition of the extended behaviors are as follows:

$$\begin{array}{rcl} B\_native & : & \texttt{T\_history}\langle\texttt{T\_collection}\langle\texttt{T\_behavior}\rangle\rangle \\ B\_inherited & : & \texttt{T\_history}\langle\texttt{T\_collection}\langle\texttt{T\_behavior}\rangle\rangle \\ B\_interface & : & \texttt{T\_history}\langle\texttt{T\_collection}\langle\texttt{T\_behavior}\rangle\rangle \end{array}$$

Each behavior now returns a collection of a collection of timestamped behaviors. Adding a new behavior to a type changes the history of the type's interface to include the new behavior. The old interface of the type is still accessible at a time before the change was made.

Note that we do not need to explicitly maintain separate histories for each of these behaviors. For example, in an implementation we can choose to only maintain the native behaviors of a type.

The entire interface of a type can be derived by unioning the native behaviors of all the supertypes of the type. The inherited behaviors can be derived by taking the difference of the interface and the native behaviors of the type. As another alternative, we may choose to maintain the interface of a type and derive the native and inherited behaviors. In this approach, the native behaviors of a type can be derived by unioning the interfaces of the direct supertypes and subtracting the result from the interface of the type. The inherited behaviors can be derived in the same way as above.

With the time-varying interface extensions, we can determine the various aspects of a type's interface at any time of interest. For example, Figure 3 shows the history of the entire interface for the type `T_person`.

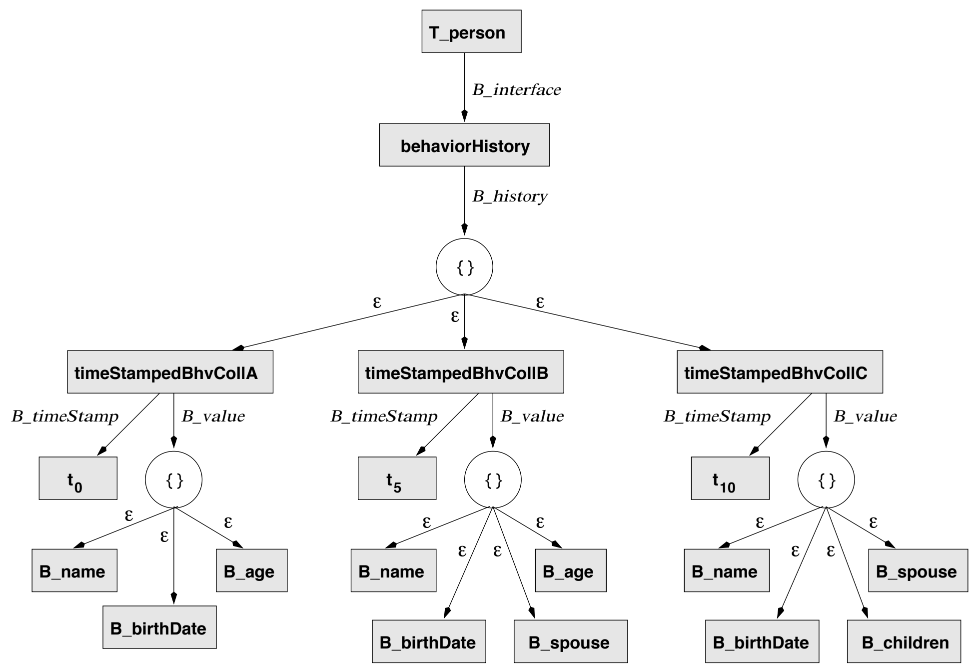

Figure 3: Interface history of type `T_person`.

At time t_0, behaviors *B_name*, *B_birthDate*, and *B_age* are defined on `T_person` and the initial history of `T_person`'s interface is $\{< t_0, \{B\_name, B\_birthDate, B\_age\} >\}$. At time t_5, behavior *B_spouse* is added to `T_person`. To reflect this change, the interface history is updated to $\{< t_0, \{B\_name, B\_birthDate, B\_age\} >, < t_5, \{B\_name, B\_birthDate, B\_age, B\_spouse\} >\}$. This shows that between t_0 and t_5 only behaviors *B_name*, *B_birthDate*, and *B_age* are defined and at t_5 behaviors *B_name*, *B_birthDate*, *B_age*, *B_spouse* exist. Next, at time t_{10}, behavior *B_age* is dropped from type `T_person` and at the same time behavior *B_children* is added. The final history of the interface of `T_person` after this change is $\{<t_0, \{B\_name, B\_birthDate, B\_age\}>, <$

$t_5, \{B\_name, B\_birthDate, B\_age, B\_spouse\}$>, <$t_{10}, \{B\_name, B\_birthDate, B\_spouse, B\_children\}$> }[5]. The native and inherited behaviors would contain similar histories. Using this information, we can reconstruct the interface of a type at any time of interest. For example, at time t_3 the interface of type `T_person` was $\{B\_name, B\_birthDate, B\_age\}$, at time t_5 it was $\{B\_name, B\_birthDate, B\_age, B\_spouse\}$, and at time t_{10} (*now*) it is $\{B\_name, B\_birthDate, B\_spouse, B\_children\}$.

The behavioral changes to types include the **MT-AB** and **MT-DB** entries of Table 1. These changes affect various aspects of the schema and have to be properly managed to ensure consistency of the schema.

Modify Type - Add Behavior (MT-AB). This change adds a native behavior b to a type T at time t. The **MT-AB** change has the following effects:

- The histories of the native and interface behaviors of type T need to be updated. The behavior applications $T.B\_native.B\_insert(b, t)$ and $T.B\_interface.B\_insert(b, t)$ perform this update. For example, the behavior application `T_person`.*B_interface*.*B_insert*(*B_spouse*,t_5) updates the interface history of `T_person` when behavior *B_spouse* is added to `T_person` at time t_5.
- The implementation history of behavior b needs to be updated to associate it with some function f. This is achieved by the behavior application $b.B\_implementation.B\_insert(f, t)$ (details on implementation histories of behaviors are given in Section 4.3). For example, if the function associated with behavior *B_spouse* is the stored function s_{spouse}, then the implementation history of *B_spouse* is updated using the behavior application *B_spouse*.*B_implementation*.*B_insert*(s_{spouse},t_5).
- The history of inherited and interface behaviors of all subtypes of type T needs to be adjusted. That is,

  $\forall T' \mid T'$ subtype-of $T, T'.B\_inherited.B\_insert(b, t)$ and $T'.B\_interface.B\_insert(b, t)$

  For example, the histories of inherited and interface behaviors of types `T_employee` and `T_patient` (see Figure 1) need to be adjusted to reflect the addition of behavior *B_spouse* in type `T_person` at time t_5. For the `T_employee` type, this is accomplished using the

[5]Note that in Figure 3 objects that are repeated in the timestamped collections are actually the same object. For example, the *B_name* object in all three timestamped collections is the same object. It is shown three times in the figure for clarity.

behavior applications `T_employee`.$B\_interface.B\_insert(B\_spouse,t_5)$ and `T_employee`. $B\_inherited.B\_insert(B\_spouse,t_5)$. Similar behavior applications are carried out for `T_patient`.

Modify Type - Drop Behavior (MT-DB). This change drops a native behavior b from a type T at time t. When a behavior is dropped, its native definition is propagated to the subtypes unless the behavior is inherited by the subtype through some other chain. In this way, as with the supertypes, the subtypes of a type also retain their original behaviors. Thus, only the single type involved in the operation actually drops the behavior and the overall interface of the subtypes and supertypes are not affected by the change. Many behavior inheritance semantics are possible. One such semantics is that when a native behavior is dropped from a type, all subtypes retain that behavior. This means that if another supertype of the subtype defines this behavior, there is no change. Otherwise, the behavior in the subtype moves from the inherited set to the native set. This is the semantics we are modeling in this paper. If any other behavior inheritance semantics are used, appropriate changes can easily be made to the temporal histories. The **MT-DB** change has the following effects:

- The native behaviors history of type T changes. The behavior application $T.B\_native.$ $B\_remove(b, t)$ performs this update. For example, the behavior application `T_person`. $B\_native.B\_remove(B\_age,t_{10})$ updates the history of native behaviors of `T_person` when the behavior B_age is dropped from type `T_person`.
- The native and inherited behavior histories of the subtypes of T (possibly) change. For example, the behavior applications `T_employee`.$B\_native.B\_insert(B\_age,t_{10})$ and `T_employee`.$B\_inherited.B\_remove(B\_age,t_{10})$ add behavior B_age to the native behaviors of `T_employee`, and drop behavior B_age from the inherited behaviors of `T_employee` respectively, when B_age is dropped from `T_person` at t_{10}. This is because B_age is not inherited by `T_employee` through any other chain. If B_age was inherited by `T_employee` from some other supertype, nothing would change. Similar behavior applications are carried out for type `T_patient`.

4.3 Changing Implementations of Behaviors

Each behavior defined on a type has a particular implementation for that type. The $B\_implementation$ behavior defined on `T_behavior` is applied to a behavior, accepts a type as an argument and returns the implementation (function) of the receiver behavior for the given type. In order to model the aspect of schema evolution that deals with changing the implementations of behaviors on types, we maintain a history of implementation changes by extending the *B_implementation* behavior with time-varying properties. The definition of the extended behavior is as follows:

$$B\_implementation : \texttt{T\_type} \rightarrow \texttt{T\_history}\langle\texttt{T\_function}\rangle$$

With this behavior we can determine the implementation of a behavior defined on a type at any time of interest. For example, Figure 4 shows the history of the implementations for behavior *B_age* on type `T_person`. There are two kinds of implementations for behaviors [Pet94]. A *computed function* consists of runtime calls to executable code and a *stored function* is a reference to an existing object in the object database.

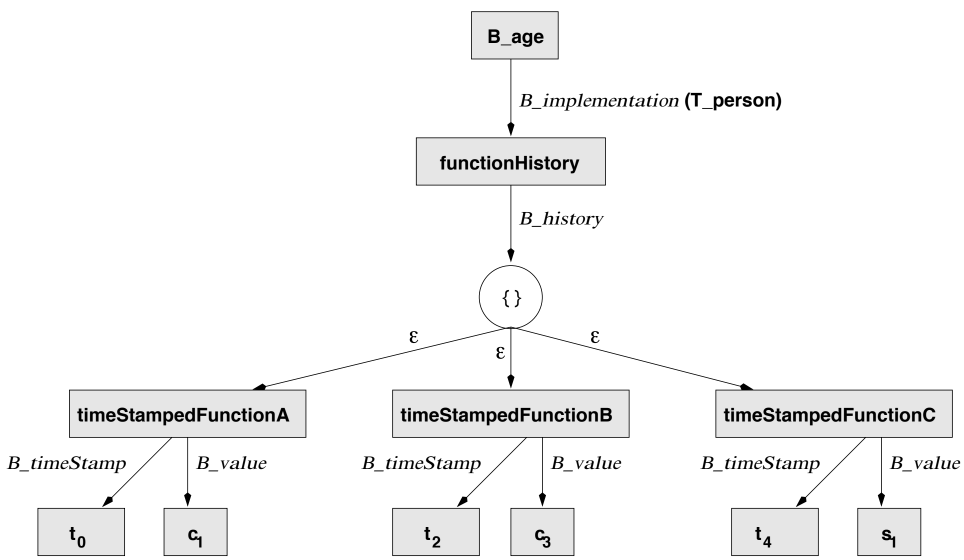

Figure 4: Implementation history of behavior *B_age* on type `T_person`.

In Figure 4, we use c_i to denote a computed function, s_i to denote a stored function. At time t_2, the implementation of *B_age* changed from the computed function c_1 to the computed function c_3. At time t_4, the implementation of *B_age* changed from the computed function c_3 to the stored function s_1. All these changes are reflected in the implementation history of behavior *B_age*, which is $\{<t_0, c_1>, <t_2, c_3>, <t_4, s_1>\}$.

Using the results of this section and Section 4.2, we can reconstruct the behaviors, their implementations and the object representations[6] for any type at any time t. For example, the interface

[6]Stored functions associated with behaviors allow us to reconstruct object representations (i.e., states of objects)

of type T_person at time t_3 is given by the behavior application T_person.$[t_3]$*B_interface* which results in $\{B\_name, B\_birthDate, B\_age\}$, as shown in Figure 3. We use the syntax $o.[t]b$ to denote the application of behavior b to object o at time t. The implementation of *B_age* at time t_3 is given by *B_age*.$[t_3]$*B_implementation*(T_person) which is c_3, as shown in Figure 4.

In this paper we are assuming that there is no implementation inheritance. That is, if the binding of a behavior to a function changes in a type, the bindings of that behavior in the subtypes are unaffected. If implementation inheritance is desired, it can easily be modeled by temporal histories similarly to behavioral inheritance.

4.4 Changing Subtype/Supertypes of a Type

In Section 4.2 we described how the changes in a type's interface was one aspect in which a type evolves. Another aspect of a type that can change over time is the relationships between types. These include adding a direct supertype link and dropping a direct supertype link. The *B_supertypes* and *B_subtypes* behaviors defined on T_type return the direct supertypes and subtypes of the receiver type, respectively. In order to model the structure of the type lattice through time, we define histories of supertype and subtype changes of a type by extending the *B_supertypes* and *B_subtypes* behaviors with time-varying properties:

$$\begin{aligned} B\_supertypes &: \texttt{T\_history}\langle\texttt{T\_collection}\langle\texttt{T\_type}\rangle\rangle \\ B\_subtypes &: \texttt{T\_history}\langle\texttt{T\_collection}\langle\texttt{T\_type}\rangle\rangle \end{aligned}$$

Using the *B_supertypes* and *B_subtypes* behaviors, we can reconstruct the structure of a type's supertype and subtype lattice at any time of interest. To facilitate this, the derived behaviors *B_superlattice* and *B_sublattice* are defined on T_type:

$$\begin{aligned} B\_superlattice &: \texttt{T\_history}\langle\texttt{T\_poset}\langle\texttt{T\_type}\rangle\rangle \\ B\_sublattice &: \texttt{T\_history}\langle\texttt{T\_poset}\langle\texttt{T\_type}\rangle\rangle \end{aligned}$$

The behavior *B_superlattice* is derived by recursively applying *B_supertypes* until T_object is reached, while the behavior *B_sublattice* is derived by recursively applying *B_subtypes* until T_null is reached. In both cases, the intermediate results are partially ordered. Figure 5 shows the supertype lattice history for type T_employee.

for any type at any time t. This is useful in propagating changes to the underlying object instances. In this paper however, we are concerned primarily with the effects of schema changes on the schema itself.

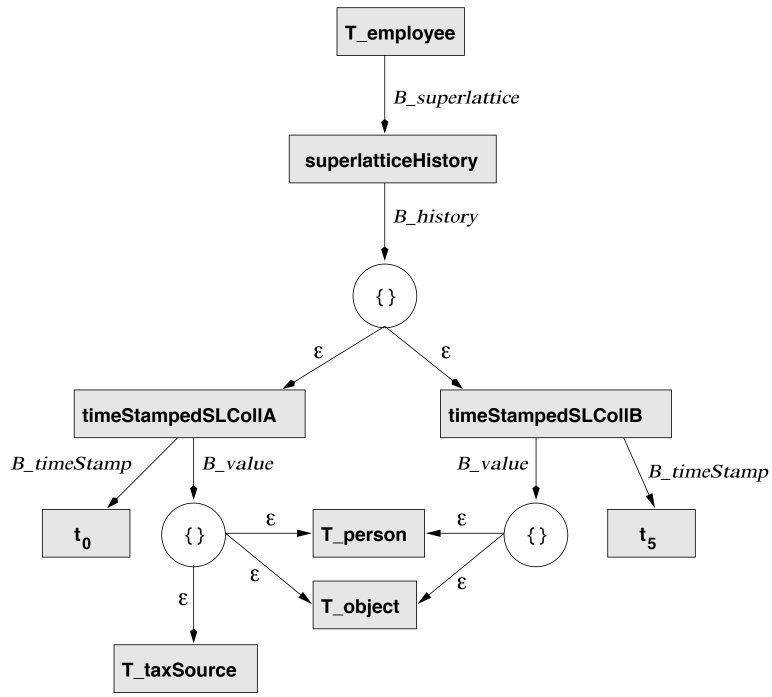

Figure 5: Supertype lattice history for type `T_employee`.

At time t_0, the superlattice history of type `T_employee` included the types `T_person`, `T_taxSource`, and `T_object`. At time t_5, the supertype link between `T_employee` and `T_taxSource` is dropped. To reflect this change, the superlattice history of `T_employee` is updated to $\{<t_0, \{$`T_person`, `T_taxSource`, `T_object`$\}>, <t_5, \{$`T_person`, `T_object`$\}>\}$.

The relationships between types include the **MT-ASL** and **MT-DSL** entries of Table 1. Similar to the behavioral changes to types discussed in Section 4.2, the relationships between types affect various aspects of the schema and have to be properly managed to ensure consistency of the schema.

Modify Type - Add Supertype Link (MT-ASL). Add a type, say S, as a direct supertype of another type, say T at time t. The **MT-ASL** change has the following effects:

- The history of the collection of supertypes of type T is updated. The behavior application $T.B\_supertypes.B\_insert(S,t)$ performs this update. The history of the superlattice of T is adjusted accordingly. For example, adding the supertype link between `T_employee` and `T_taxSource` at t_0 necessitates an update to the history of supertypes for `T_employee`. This is done by the behavior application `T_employee`.*B_supertypes*.*B_insert* (`T_taxSource`,t_0). The history of the direct supertypes of `T_employee` would then be $\{<t_0, \{$`T_taxSource`$\}>\}$.

- The history of the collection of subtypes of type S is updated. The behavior application $S.B\_subtypes.B\_insert(T, t)$ performs this update. The history of the sublattice of S is adjusted accordingly. In this case, the history of the collection of subtypes of `T_taxSource` has to be updated. This is done by the behavior application `T_taxSource`.*B_subtypes*.*B_insert*(`T_employee`,t_0). The history of the direct subtypes of `T_taxSource` would then be $\{<t_0, \{$`T_employee`$\}>\}$.

- The behaviors of S are inherited by T and all the subtypes of T. Therefore, the inherited behavior history of T and all subtypes of T is adjusted. The current behaviors of S are inherited by T and all subtypes of T, and timestamped with t - the creation time of the supertype link.

  $$\forall b \in S.B\_interface.B\_history.B\_last, \forall T' \mid T' \text{ subtype-of } T, T'.B\_inherited.B\_insert(b, t)$$

  Behavior *B_last* returns the collection of behaviors that are currently valid from the interface history of S. Let us assume `T_taxSource` has the behavior *B_taxBracket* defined at t_0. *B_taxBracket* then has to be added to the history of inherited behaviors of `T_employee`. This is done by the behavior application `T_employee`.*B_inherited*.*B_insert* (*B_taxBracket*,t_0). The history of the inherited behaviors would then be $\{<t_0, \{B\_name, B\_birthDate, B\_age, B\_taxBracket\}>\}$. Behaviors *B_name*,*B_birthDate*,*B_age* are inherited from type `T_person` (see Figure 3), while behavior *B_taxBracket* is inherited from type `T_taxSource`.

Modify Type - Drop Supertype Link (MT-DSL). Drop a direct supertype link between two types (a direct supertype link to `T_object` cannot be dropped) at time t. Consider types T and S where S is the direct supertype of T. Then, removing the direct supertype link between T and S at time t has the following effects:

- Adjust the history of supertypes of T and the history of subtypes of S. For example, dropping the supertype link between `T_employee` and `T_taxSource` at t_5 requires updating the history of supertypes of `T_employee` and history of subtypes of `T_taxSource`. This is carried out using the behavior applications `T_employee`.*B_supertypes*.*B_remove*(`T_taxSource`,t_5) and `T_taxSource`.*B_subtypes*.*B_remove*(`T_employee`,t_5).

- The **MT-ASL** operation is carried out from T to every supertype of S, unless T is linked to the supertype through another chain. This operation is not required when the supertype link between `T_employee` and `T_taxSource` is dropped because `T_employee` is linked to the supertype of `T_taxSource` (`T_object`) through `T_person`.

- The **MT-ASL** operation is carried out from each subtype of T to S, unless the subtype is linked to S through another chain. This operation requires adding a supertype link between `T_null` and `T_taxSource`.
- The native behaviors of S are dropped from the interface of T. That is, the history of inherited behaviors of T is adjusted. This means the behavior *B_taxBracket*, defined natively on `T_taxSource`, has to be dropped from the history of inherited behaviors of `T_employee`. The behavior application `T_employee`.*B_inherited*.*B_remove*(*B_taxBracket*,t_5).

5 Queries

In this section we show how queries can be constructed using the TIGUKAT query language (TQL) [PLÖS93] to retrieve schema objects at any time in their evolution histories. This gives software designers a temporal user interface which provides a practical way of accessing temporal information in their experimental and incremental design phases. TQL incorporates reflective temporal access in that it can be used to retrieve both objects, and schema objects in a uniform manner. Hence, TQL does not differentiate between queries (which query objects) and meta-queries (which query schema objects).

5.1 The TIGUKAT Query Language

In this section, we briefly discuss the TIGUKAT Query Language (TQL). TQL[7] is based on the SQL paradigm [Dat87] and its semantics is defined in terms of the object calculus. Hence, every statement of the language corresponds to an equivalent object calculus expression. The basic query statement of TQL is the *select statement* which operates on a set of input collections, and returns a new collection as the result:

select < *object variable list* >
[**into** < *collection name* >]
from < *range variable list* >
[**where** < *boolean formula* >]

The *select clause* in this statement identifies the objects to be returned in a new collection. There can be one or more object variables with different formats (constant, variables, path expressions

[7]TQL was developed before the release of OQL [Cat94]. It is quite similar to OQL in structure.

or index variables) in this clause. They correspond to free variables in object calculus formulas. The *into clause* declares a reference to a new collection. If the *into clause* is not specified, a new collection is created; however, there is no reference to it. The *from clause* declares the ranges of object variables in the *select* and *where* clauses. Every object variable can range over either an existing collection, or a collection returned as a result of a subquery, where a subquery can be either given explicitly, or as a reference to a query object. The *where clause* defines a boolean formula that must be satisfied by objects returned by a query.

Having described TQL, we show in the next section how temporal objects can uniformly be queried using behavior applications without changing any of the basic constructs of TQL.

5.2 Query Examples

Example 5.1 Return the time when the behavior *B_children* was added to the type `T_person`.

select *b.B_timestamp*

from *b* **in** `T_person`.*B_interface.B_history*

where *B_children* **in** *b.B_value*

The result of this query would be the time t_{10} as seen in Figure 3. □

Example 5.2 Return the types that define behaviors *B_age* and *B_taxBracket* as part of their interface.

select *T*

from *T* **in C_type**

where (*b*1 **in** *T.B_interface.B_history* **and** *B_age* **in** *b*1*.B_value*) **or**

(*b*2 **in** *T.B_interface.B_history* **and** *B_taxBracket* **in** *b*2*.B_value*)

This query would return the types `T_person`, `T_taxSource`, `T_employee`, and `T_null`. The type `T_person` defines behavior *B_age* natively (see Figure 3), while the type `T_taxSource` defines behavior *B_taxBracket* natively. The behaviors *B_age* and *B_taxBracket* are inherited by types `T_employee` and `T_null` since they are subtypes of `T_person` and `T_taxSource` as shown in Figure 1. □

Example 5.3 Return the implementation of behavior *B_age* in type `T_person` at time t_1.

select *i.B_value*

from *i* **in** *B_age.B_implementation*(`T_person`)*.B_history*

where *i.B_timestamp.B_lessthaneqto*(t_1)

The behavior *B_lessthaneqto* is defined on type `T_timeStamp` and checks if the receiver timestamp is less than or equal to the argument timestamp. The result of the query is the computed function c_1 as shown in Figure 4. □

Example 5.4 Return the super-lattice of type `T_employee` at time t_3.

select r.*B_value*

from r **in** `T_employee`.*B_super-lattice*.*B_history*

where r.*B_timestamp*.*B_lessthaneqto*(t_3)

The super-lattice of `T_employee` at t_3 consists of the types `T_person`, `T_taxSource`, and `T_object`. This is shown in Figure 5. □

Example 5.5 Return the types that define behavior *B_age* with the same implementation as one of their supertypes.

select T

from T **in C_type**, S **in** T.*B_supertypes*.*B_history*,

   i **in** *B_age*.*B_implementation*(T).*B_history*,

   j **in** *B_age*.*B_implementation*(S.*B_value*).*B_history*

where b **in** S.*B_value*.*B_interface*.*B_history* **and** *B_age* **in** b.*B_value* **and**

   i.*B_value* = j.*B_value* **and** i.*B_timestamp* = j.*B_timestamp*

This query would return the types `T_employee`, `T_patient`, and `T_null`, assuming the implementation of behavior *B_age* is not changed when it is inherited by these types. □

6 Conclusion

In this paper a uniform treatment of schema evolution and temporal support for object database management systems (ODBMS) is presented. Schema evolution is managed by exploiting the functionality of a temporal object model. The evolution history of the interface of types, which includes the inherited and native behaviors of each type, describes the semantics of types through time. Using the interface histories the interface of a type can be reconstructed at any time of interest. The evolution histories of the supertype and subtype links of types describe the structure of the lattice through time. Using these histories, the structure of the lattice can be reconstructed at any time of interest. The implementation histories of behaviors give us the implementations of behaviors on types at any time of interest. From these, we can reconstruct the representation of objects by examining the stored functions associated with behaviors at a given time. The TIGUKAT query language gives designers a practical way of accessing temporal information in their experimental and incremental design phases.

Our next step is to give a comprehensive treatment to the change propagation problem during schema evolution. That is, devising methods to propagate schema changes to the existing object instances in the TIGUKAT temporal ODBMS. In order for the instances to remain meaningful after the schema has changed, either the relevant instances must be coerced into the new definition of the schema or a new version of the schema must be created leaving the old version intact. Conversion of objects can be optional in our model. Since the evolution history of schema objects is maintained, all the information for older objects is available and we can use this information to continue processing these objects in the old way. Since our model is time based, the old information of the object is available. Thus, even if objects are coerced to a newer schema definition, historical queries can be run by giving an appropriate time point in the history of the object.

To overcome the corrective nature of schema evolution, the concept of *schema versioning* in ODBMSs has been proposed [SZ86, SZ87, KC88, ALP91, MS92, MS93]. In most of these systems, a change to a schema object may result in a new *version* of the schema object, or the schema in general. However, schema changes are usually of a finer granularity than definable versions. This implies that not every schema change should necessarily result in a new version. Rather, one should be able to define a version during any stable period in the evolutionary history of the schema. Within a particular version, the evolution of the schema should be traceable. For example, in an engineering design application many components of an overall design may go through several modifications in order to produce a final product. Furthermore, each intermediate version of the component may have certain properties that need to be retained as a historical record of that particular component (the different versions may have been used in other products). The interconnection of the various versions of components also gives rise to versions of the overall design. The resulting designs may be part of others and so on. Our contention is that schema evolution using temporal modeling sets the stage for full-fledged version control. We intend to use the schema evolution policies reported in this paper as a basis for version control in ODBMSs.

References

[ALP91] J. Andany, M. Leonard, and C. Palisser. Management of Schema Evolution in Databases. In *Proc. 17th Int'l Conf. on Very Large Data bases*, pages 161–170, September 1991.

[Ari91] G. Ariav. Temporally oriented data definitions: Managing schema evolution in temporally oriented databases. *Data & Knowledge Engineering*, (6):451–467, 1991.

[BFG96] E. Bertino, E. Ferrari, and G. Guerrini. A Formal Temporal Object-Oriented Data Model. In *Proc. 5th Int'l Conf. on Extending Database Technology*, March 1996.

[BKKK87] J. Banerjee, W. Kim, H-J. Kim, and H.F. Korth. Semantics and Implementation of Schema Evolution in Object-Oriented Databases. In *Proc. ACM SIGMOD Int'l. Conf. on Management of Data*, pages 311–322, May 1987.

[Cat94] R. Cattell. *The Object Database Standard: ODMG-93.* Morgan Kaufmann Publishers, 1994.

[CITB92] W.W. Chu, I.T. Ieong, R.K. Taira, and C.M. Breant. A Temporal Evolutionary Object-Oriented Data Model and Its Query Language for Medical Image Management. In *Proc. 18th Int'l Conf. on Very Large Data Bases*, pages 53–64, August 1992.

[CPP95] C. Combi, F. Pinciroli, and G. Pozzi. Managing Different Time Granularities of Clinical Information by an Interval-Based Temporal Data Model. *Methods of Information in Medicine*, 34(5):458–474, 1995.

[Dat87] C.J. Date. *A Guide to SQL Standard.* Addison Wesley, 1987.

[DM94] A.K. Das and M.A. Musen. A Temporal Query System for Protocol-Directed Decision Support. *Methods of Information in Medicine*, 33:358–370, 1994.

[GLÖS95] I.A. Goralwalla, Y. Leontiev, M.T. Özsu, and D. Szafron. A Uniform Behavioral Temporal Object Model. Technical Report TR-95-13, University of Alberta, May 1995.

[GLÖS96] I.A. Goralwalla, Y. Leontiev, M.T. Özsu, and D. Szafron. Modeling Time: Back to Basics. Technical Report TR-96-03, University of Alberta, February 1996.

[GÖS97] I.A. Goralwalla, M.T. Özsu, and D. Szafron. Modeling Medical Trials in Pharmacoeconomics using a Temporal Object Model. *Computers in Biology and Medicine - Special Issue on Time-Oriented Systems in Medicine*, 1997. In Press.

[GTC+90] S. Gibbs, D.C. Tsichritzis, E. Casais, O.M. Nierstrasz, and X. Pintado. Class Management for Software Communities. *Communications of the ACM*, 33(9):90–103, September 1990.

[KBCG90] W. Kim, J. Banerjee, H.T. Chou, and J.F. Garza. Object-oriented database support for CAD. *Computer Aided Design*, 22(8):469–479, 1990.

[KC88] W. Kim and H-J. Chou. Versions of Schema for Object-Oriented Databases. In *Proc. 14th Int'l Conf. on Very Large Data Bases*, pages 148–159, 1988.

[KFT91] M.G. Kahn, L.M. Fagan, and S. Tu. Extensions to the Time-Oriented Database Model to Support Temporal Reasoning in Medical Expert Systems. *Methods of Information in Medicine*, 30:4–14, 1991.

[KS92] W. Kafer and H. Schoning. Realizing a Temporal Complex-Object Data Model. In *Proc. ACM SIGMOD Int'l. Conf. on Management of Data*, pages 266–275, 1992.

[LGÖS97] J.Z. Li, I.A. Goralwalla, M.T. Özsu, and Duane Szafron. Modeling Video Temporal Relationships in an Object Database Management System. In *Proceedings of Multimedia Computing and Networking (MMCN97)*, February 1997.

[MS90] Edwin McKenzie and Richard Snodgrass. Schema evolution and the relational algebra. *Information Systems*, 15(2):207–232, 1990.

[MS92] S.R. Monk and I. Sommerville. A Model for Versioning of Classes in Object-Oriented Databases . In *10th British National Conference on Databases (BNCOD '92), Aberdeen, Scotland July 1992*, pages 42–58, July 1992.

[MS93] Simon Monk and Ian Sommerville. Schema Evolution in OODBs using Class Versioning. *ACM SIGMOD Record*, 22(3):16–22, September 1993.

[NR89] G.T. Nguyen and D. Rieu. Schema evolution in object-oriented database systems. *Data & Knowledge Engineering*, 4:43–67, 1989.

[ÖPS+95] M.T. Özsu, R.J. Peters, D. Szafron, B. Irani, A. Lipka, and A. Munoz. TIGUKAT: A Uniform Behavioral Objectbase Management System. *The VLDB Journal*, 4:100–147, August 1995.

[Pet94] R.J. Peters. *TIGUKAT: A Uniform Behavioral Objectbase Management System*. PhD thesis, University of Alberta, 1994.

[PLÖS93] R.J. Peters, A. Lipka, M.T. Özsu, and D. Szafron. An Extensible Query Model and Its Languages for a Uniform Behavioral Object Management System. In *Proc. Second Int'l. Conf. on Information and Knowledge Management*, November 1993.

[PÖ93] R.J. Peters and M.T. Özsu. Reflection in a Uniform Behavioral Object Model. In *Proc. 12th Int'l Conf. on the Entity Relationship Approach*, pages 37–49, December 1993.

[PÖ97] R.J. Peters and M.T. Özsu. An Axiomatic Model of Dynamic Schema Evolution in Objectbase Systems. *ACM Transactions on Database Systems*, 22(1):75–114, March 1997.

[PS87] D.J. Penney and J. Stein. Class Modification in the GemStone Object-Oriented DBMS. In *Proc. of the Int'l Conf on Object-Oriented Programming: Systems, Languages, and Applications*, pages 111–117, October 1987.

[Rod91] J.F. Roddick. Dynamically Changing Schemas within Database Models. *Australian Computer Journal*, 23(3):105–109, 1991.

[Rod92] J.F. Roddick. SQL/SE- A Query Language Extension for Databases Supporting Schema Evolution. *ACM SIGMOD Record*, 21(3):10–16, 1992.

[Rod95] J.F. Roddick. A Survey of Schema Versioning Issues for Database Systems. *Information and Software Technology*, 37(7):383–393, 1995.

[RS91] E. Rose and A. Segev. TOODM - A Temporal Object-Oriented Data Model with Temporal Constraints. In *Proc. 10th Int'l Conf. on the Entity Relationship Approach*, pages 205–229, October 1991.

[SC91] S.Y.W. Su and H.M. Chen. A Temporal Knowledge Representation Model OSAM*/T and its Query Language OQL/T. In *Proc. 17th Int'l Conf. on Very Large Data bases*, 1991.

[Sjø93] Dag Sjøberg. Quantifying Schema Evolution. *Information and Software Technology*, 35(1):35–44, January 1993.

[Sno95] R. Snodgrass. *The TSQL2 Temporal Query Language*. Kluwer Academic Publishers, 1995.

[SS77] J.M. Smith and D.C.P. Smith. Database Abstractions: Aggregation and Generalization. *ACM Transactions on Database Systems*, 2(2):105–133, 1977.

[SZ86] A.H. Skarra and S.B. Zdonik. The Management of Changing Types in an Object-Oriented Database. In *Proc. of the Int'l Conf on Object-Oriented Programming: Systems, Languages, and Applications*, pages 483–495, September 1986.

[SZ87] A.H. Skarra and S.B. Zdonik. Type Evolution in an Object-Oriented Database. In *Research Directions in Object-Oriented Programming*, pages 393–415. M.I.T. Press, 1987.

[WD92] G. Wuu and U. Dayal. A Uniform Model for Temporal Object-Oriented Databases. In *Proc. 8th Int'l. Conf. on Data Engineering*, pages 584–593, February 1992.