Data Models with Multiple Temporal Dimensions: Completing the Picture

Carlo Combi and Angelo Montanari

Department of Mathematics and Computer Science

University of Udine, Italy

{*combi,montana*}*@dimi.uniud.it t*

Talk Overview

- Introduction;
- The basic temporal dimensions: valid and transaction times;
- A third temporal dimension: the event time;
- One event time is not enough;
- A new temporal dimension: the availability time;
- Conclusions and further work.

State of the art

- Fundamental temporal dimensions: valid and transaction times.
- However, meaningful temporal aspects of facts cannot be naturally modeled by means of valid and transaction times. As an example, they do not allow one to distinguish between retroactive and delayed updates.
- A third temporal dimension: *event time*
  - it makes it possible to model retroactive, on-time, and proactive updates
  - the combined use of valid, transaction, and event times allows one to maintain different past states generated by retroactive and proactive updates as well as by error corrections and delayed updates

Original contributions

- We first refine the notion of *event time*: we associate with each fact both the occurrence time of the event that initiates its validity interval and the occurrence time of the event that terminates it.
- Then, we introduce a further temporal dimension (*availability time*), which can be viewed as the information system counterpart of the real-world event time.

Valid and Transaction Times

Valid time: the *valid time* (VT) of a fact is the time when the fact is true in the modeled reality.

Transaction time: the *transaction time* (TT) of a fact is the time when the fact is current in the database and may be retrieved.

- Valid time is usually provided by database users, while transaction time is system-generated and supplied.
- Valid and transaction times are orthogonal dimensions: each of them can be independently recorded or not and has specific properties.
- Valid and transaction times can be related to each other in different ways. Jensen and Snodgrass propose several classifications (*specializations*) of (bi)temporal relations, based on the relationships between valid and transaction times of timestamped facts.

A concrete situation

Example: on August 10, 1998, the physician prescribes a bipuvac-based therapy from 10:00 to 14:00. Data about the therapy is entered into the database at 9:00. Due to the unexpected evolution of the patient state, the bipuvac infusion is stopped at 11:15 and replaced by a diazepam-based therapy from 11:25 to 14:00. The new facts are entered at 12:00. (Bipuvac and diazepam are drugs commonly used in anesthesia.)

Drug	VT	TT
bipuvac	[98Aug10;10:00, 98Aug10;14:00)	[98Aug10;9:00, 98Aug10;12:00)
bipuvac	[98Aug10;10:00, 98Aug10;11:15)	[98Aug10;12:00, ∞)
diazepam	[98Aug10;11:25, 98Aug10;14:00)	[98Aug10;12:00, ∞)

Some issues

Problem: can a single valid time be always associated with a fact? *One valid time is enough (the problem is that of properly identifying, at the conceptual level, what constitutes a fact.)*

Linear vs branching time: we assume the valid time domain to be linearly ordered, thus excluding branching valid times. Branching valid time is needed to model situations where different perceptions of the reality exist (one for each distinct timeline).

Event Time

Event time: the *event time* (ET) of a fact is the occurrence time of the real-world event that generates the fact.

Examples:

- an on-time promotion event, that increases the rank of a physician, occurs on October 1, 1997, but its effects are recorded in the database on November 1 (delayed update);
- a retroactive promotion event, whose effects are immediately recorded in the database, occurs on November 1, but the increase in rank of the physician becomes valid since October 1 (retroactive update).

Valid and Event Times

- On-time events: the validity interval starts at the occurrence time of the event (e.g., hospitalization starts immediately after the family doctor decision).
- Retroactive events: the validity interval starts before the occurrence time of the event (e.g., on December 21, 1997, the manager of the hospital decided an increase of 10% of the salary of the physicians of the Pathology Department, starting from December 1, 1997).
- Proactive events: the validity interval starts after the occurrence time of the event (e.g., on January 29, 1998, the family doctor decides the patient hospitalization on February 15, 1998).

Transaction and Event Times

- On-time update: the transaction time coincides with the event time. This situation happens when data values are inserted in the database as soon as they are generated.
- Delayed update: the transaction time is greater than the event time. This is the case when data values are inserted some time after their generation.
- Anticipated update: the transaction time is less than the event time. This is the case when data values are entered into the database before the occurrence time of the event that generates them. Such a notion of anticipated update is useful to model hypothetical courses of events.

One Event Time is not enough

IF there are gaps in temporal validity (a relation modeling admissions to hospital, which only records information about inpatients),

OR only incomplete information about the effects of an event is available (we know that an event that changes the rank of the physician occurred, but we do not know if it is a promotion or a downgrading event),

OR the expected termination of a validity interval is (must be) revised, while its initiation remains unchanged (a prescribed therapy needs to be stopped due to an unexpected evolution of the patient state),

THEN we need to distinguish between the occurrence times of the two events that respectively initiate and terminate the validity interval of the fact.

A concrete situation

Example: on August 10, 1998, at 8:00, the physician prescribes a bipuvac-based therapy from 10:00 to 14:00. Data about the therapy is entered into the database at 9:00. Due to the unexpected evolution of the patient state, at 11:00 the physician decides a change in the patient therapy. Accordingly, the bipuvac infusion is stopped at 11:15 and replaced by a diazepam-based therapy from 11:25 to 14:00. The new facts are entered at 12:00.

Two (incorrect) solutions

Drug	VT	ET	TT
bipuvac	[98Aug10;10:00, 98Aug10;14:00)	98Aug10;8:00	[98Aug10;9:00, 98Aug10;12:00)
bipuvac	[98Aug10;10:00, 98Aug10;11:15)	98Aug10;8:00	[98Aug10;12:00, ∞)
diazepam	[98Aug10;11:25, 98Aug10;14:00)	98Aug10;11:00	[98Aug10;12:00, ∞)

Drug	VT	ET	TT
bipuvac	[98Aug10;10:00, 98Aug10;14:00)	98Aug10;8:00	[98Aug10;9:00, 98Aug10;12:00)
bipuvac	[98Aug10;10:00, 98Aug10;11:15)	98Aug10;11:00	[98Aug10;12:00, ∞)
diazepam	[98Aug10;11:25, 98Aug10;14:00)	98Aug10;11:00	[98Aug10;12:00, ∞)

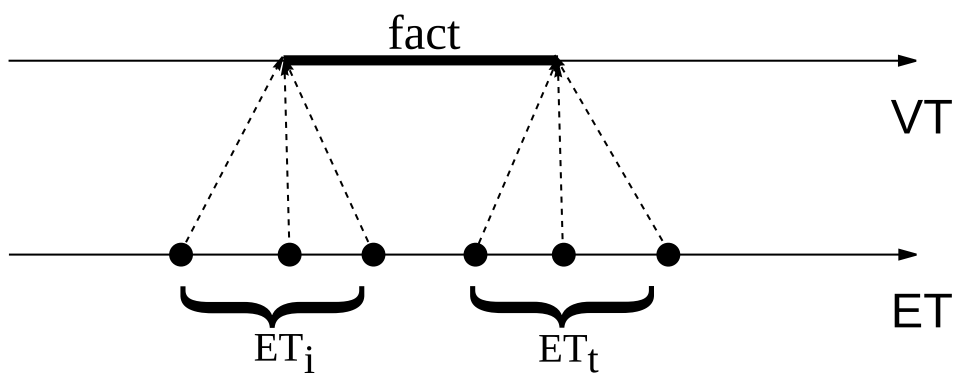
Event time revisited

Event time: the *event time* of a fact is the occurrence time of a real-world event that either initiates or terminates the validity interval of the fact.

Drug	VT	ET_i	ET_t	TT
bipuvac	[98Aug10;10:00, 98Aug10;14:00)	98Aug10;8:00	98Aug10;8:00	[98Aug10;9:00, 98Aug10;12:00)
bipuvac	[98Aug10;10:00, 98Aug10;11:15)	98Aug10;8:00	98Aug10;11:00	[98Aug10;12:00, ∞)
diazepam	[98Aug10;11:25, 98Aug10;14:00)	98Aug10;11:00	98Aug10;11:00	[98Aug10;12:00, ∞)

Some remarks

Valid and event time relationships:



This distinction between initiating and terminating event times comes from ideas underlying classical formalisms in the area of **reasoning about actions and change**.

The choice of adding **event time(s) as a separate temporal dimension** has been extensively debated in temporal databases. The basic issue is whether or not events and facts must be dealt with in a different way.

Time and Information Systems

By **information system** we mean the set of information flows of an organization and the human and computer-based resources that manage them.

From such a point of view, we may need to model **the time at which** (someone/something within) **the information system becomes aware of a fact** as well as the time at which the fact is stored into the database. While the latter temporal aspect is captured by the transaction time, the former has never been explicitly modeled.

A concrete situation with an incorrect solution

Example: Due to a trauma that occurred on September 15, 1997, Mary suffered from a severe headache starting from October 1. On October 7, Mary was visited by a physician. On October 9, the physician administered her a suitable drug. The day after, the physician entered acquired information about Mary's medical history into the database. On October 15, the patient told the physician that her headache stopped on October 14; the physician entered this data into the database on the same day.

symptom	VT	ET_i	ET_t	TT
headache	[97Oct1, ∞)	97Sept15	null	[97Oct10, ∞)

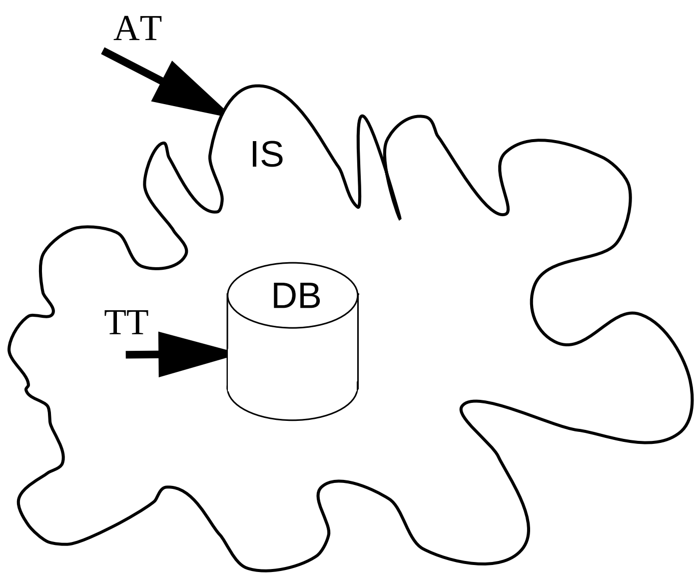
Database instance after the first update.

symptom	VT	ET_i	ET_t	TT
headache	[97Oct1, ∞)	97Sept15	null	[97Oct10, 97Oct15)
headache	[97Oct1, 97Oct14)	97Sept15	97OCt9	[97Oct15, ∞)

Database instance after the second update.

The notion of Availability Time

In general, the time at which information becomes available precedes, but not necessarily coincides with, the time at which it is recorded in the database, and thus we cannot use transaction time to model it.



Availability (AT) and transaction (TT) times and their relationships with the database (DB) and information systems (IS).

A preliminary definition

In many applications, where data insertions are grouped and executed in batches, possibly on the basis of previously filled report forms, transaction time cannot be safely taken as the time when a fact has been acquired by the information system.

In some cases, it may happen that the order according to which facts are known by the information system differs from the order in which they are stored into the database.

In many application domains, including the medical one, decisions are taken on the basis of the available information, no matter whether or not it is stored in the database.

We introduce the new temporal dimension *availability time* representing the time at which a fact becomes available to the information system.

The correct solution

The availability time allows us to completely model the previous example: the availability time of the first tuple, which is the time at which the physician learns about Mary's headache onset (October 7), strictly precedes the beginning of the transaction time interval (October 10), while the database time and the beginning of the transaction time interval of the second tuple coincide (October 15).

symptom	VT	ET_i	ET_t	AT	TT
headache	[97Oct1, ∞)	97Sept15	null	97Oct7	[97Oct10, 97Oct15)
headache	[97Oct1, 97Oct14)	97Sept15	97Oct9	97Oct15	[97Oct15, ∞)

Database instance after the second update (revisited).

Availability and Decision Times

The concept of availability time allows us to clarify the relationships between Kim and Chakravarthy's (initiating) event time and the related notion of decision time. The ***decision time*** *of a fact is the time at which the fact is decided in the application domain of discourse. More precisely, it can be defined as the occurrence time of a real-world event, whose happening induces the decision of inserting a fact into the database. In the conceptual framework we propose, the decision time models the temporal aspects of an event, initiating the validity interval of a fact, that occurs within the information system, and thus is immediately known by it. In these circumstances, the initiating event time and the availability time coincide (and are before than or equal to the starting point of the transaction time interval).*

A problem with the Availability Time

Up to now, we did not consider the possibility that the information system acquires erroneous data. The notion of availability time can be generalized to take into account such a possibility.

Example: due to an insertion mistake (or to an imprecision in Mary's talk), the trauma has been registered as happened on September 5, 1997. Only on October 20, the mistake was discovered. The day after, the physician entered the correct data into the database.

symptom	VT	ET_i	ET_t	AT	TT
headache	[97Oct1, ∞)	97Sept5	null	97Oct7	[97Oct10, 97Oct15)
headache	[97Oct1, 97Oct14)	97Sept5	97Oct9	97Oct15	[97Oct15, 97Oct21)
headache	[97Oct1, 97Oct14)	97Sept15	97Oct9	97Oct20	[97Oct21, ∞)

Availability Time

Availability Time: the *availability time* (AT) of a fact is the time interval during which the fact is known and believed correct by the information system.

symptom	VT	ET_i	ET_t	AT	TT
headache	[97Oct1, ∞)	97Sept5	null	[97Oct7, 97Oct15)	[97Oct10, 97Oct15)
headache	[97Oct1, 97Oct14)	97Sept5	97Oct9	[97Oct15, 97Oct20)	[97Oct15, 97Oct21)
headache	[97Oct1, 97Oct14)	97Sept15	97Oct9	[97Oct20, ∞)	[97Oct21, ∞)

Transaction and Availability Times

Notice that the availability time can be viewed as ***the transaction time of the information system****. If the information system outside the database is considered a database (in general, this is not case; in our medical examples, for instance, the information system includes both databases and physicians), then the availability time of a fact is when the fact is inserted into (deleted from) the information system.*

This parallel between availability time and transaction time makes it immediately clear why the availability time has to be an interval.

This point of view on the availability time resembles the notion of temporal generalization which allows several levels, and thus several transaction times, in a database.

The complete picture

Modeled reality	***Information system***
valid time	transaction time
$\uparrow$	$\uparrow$
event times	availability time

- Transaction time is append-only.
- Valid and event times, being related to times of the represented real world, can be located either in the past or in the future, and they can be modified freely.

The complete picture (cont'd)

- The availability time is append-only in its nature, because facts previously known and believed correct by the information system cannot be changed.

  However, from the point of view of the database system, availability time would be really append-only only if there were no errors in data entry.

  Since we cannot exclude such a possibility, previous states of the information system, according to the availability time, can be revised by entering new data.

  Furthermore, even assuming data entry without errors, database and transaction times may "append" facts according to two different orders.

Conclusions

In this paper, we proposed a new, fully symmetric **conceptual data model** with multiple temporal dimensions, that revises and extends existing temporal models.

We first described a **refinement of the concept of event time**, that replaces the single event time by a pair of *initiating* and *terminating event times*.

Then, we introduced **a new temporal dimension** (*availability time*), which captures the time interval during which a fact is known and believed correct by the information system.

The availability time can be exploited, for example, to analyze the **quality of decision making in information systems** where data insertion is performed in batches and thus some delay is possible between data availability and data insertion.

Further Work

We are working at a generalization of the **temporal logic reconstruction** of valid and transaction temporal databases, that copes with both event and availability times.

Another problem we are dealing with is the identification and characterization of suitable **temporal normal forms**. Many temporal dependencies among data may indeed arise when relations are provided with multiple temporal dimensions.

Finally, suitable **indexing techniques** are needed to effectively manage temporal databases with four temporal dimensions. To this end, we are currently analyzing existing indexing techniques for multidimensional data. The existence of general constraints among the different temporal dimensions hints at the possibility of tailoring existing indexing techniques and search algorithms to our specific context.