Ambiguities in Medical Bitemporalized Relational Databases: a Referent Tracking View

Adrien BARTON[a,1], Christina KHNAISSER[a], Luc LAVOIE[a]
and Jean-François ETHIER[a,1]
[a] *GRIIS, Université de Sherbrooke, Québec, Canada*

Abstract. An analysis based on referent tracking systems shows that a classical medical application of bitemporalized relational database that uses classes as attribute values is ambiguous in several respects. This suggests that to avoid such ambiguities, bitemporalized relational databases could be structured on the basis of ontological representations that give adequate attention to particulars.

Keywords. Temporalized relational database, Referent tracking, Information ambiguity, Valid time, Transaction time

1. Introduction

Temporalized relational databases ("T-RDB" for short) are used in the medical domain to describe the temporal dimensions of relevant medical entities. To be able to exchange reliably the information stored in T-RDB with other information systems, T-RDB should be structured non-ambiguously. However, this article will show that a classical medical application of T-RDB can be ambiguous in several respects, because of its use of classes as values in an attribute. The analysis will rest on the referent tracking paradigm [1] that uses biomedical ontologies in the description of particular entities such as a given health care professional, a given patient, his diseases or pathological processes, as well as the relevant temporal dimensions.

2. Temporalized relational databases

2.1. Valid time and agent-relativization

Consider the following scenario (called "scenario A"), inspired by [2]: on April 1st 2017, Mr. Hubbard conveys to Dr. Jones that after coming back from a trip to India in January, he had high fever and nausea from February 3rd to February 19th; Dr. Jones enters into his database his diagnosis that Hubbard had malaria during that time. On May 17th, in

---

[1] Corresponding Authors; E-mail: ethierj@gmail.com; adrien.barton@gmail.com. AB acknowledges financial support by the "bourse de fellowship du département de médecine de l'université de Sherbrooke" and the CIHR-funded "Quebec SPOR Support Unit".

light of medical tests and a finer estimation of when the symptoms stopped, Dr. Jones diagnoses that Hubbard had dengue fever from February 3rd to February 24th; he corrects immediately the database.

Suppose that Jones was using a T-RDB with the valid-time relvar [3] R^V = <PAT, DIS, @V, AG>, that represents the following predicate p^V(*PAT*,*DIS*,*@V*,*AG*):

> '(*PAT* is a particular patient) and (*DIS* is a class of diseases) and (*@V* is a time) and (*AG* is a particular agent) and [*AG* currently believes the following proposition: (*PAT* has a disease of class *DIS* during *@V*)]'

where *@V* is called a "valid time". Following the realist methodology [4], this predicate distinguishes clearly the particulars from the universals or classes ("a time" refers to any mereological sum [5] of particular time instants and intervals). Note that here, the values of the attribute *DIS* are classes: this will be the cause of the problems mentioned later. On April 1st, the relation R^V includes the tuple tp_1^V = (Hubbard, Malaria, [Feb3:Feb19], Jones) (dropping the reference to 2017). On May 17th, tp_1^V is replaced by tp_2^V = (Hubbard, Dengue, [Feb3:Feb24], Jones).

2.2. Transaction time

It is important to keep track of those tuple modifications in the T-RDB for audit purposes. Therefore, the database may include the "log" [3] of relation R^V, which is itself a relvar named here "$R^{V,T}$": a bitemporal relvar including both valid times and transaction times, stating when some tuple was present in the relation R^V. Its value includes the following tuples $tp_1^{V,T}$ and $tp_2^{V,T}$ on May 17th:

	PAT	*DIS*	*@V*	*AG*	*@T*
$tp_1^{V,T}$	Hubbard	Malaria	[Feb3:Feb19]	Jones	[Apr1:May16]
$tp_2^{V,T}$	Hubbard	Dengue	[Feb3:Feb24]	Jones	[May17:ufn]

where *@T* is called a "transaction time" and "ufn" stands for "until further notice", meaning that the described tuple is still presently unchanged in R^V. A database system using valid time and transaction time is called a "bitemporalized RDB" ("B-RDB").

$tp_1^{V,T}$ means that according to R^V, Jones believed only from April 1st to May 16th (because of the closed-world assumption) that Hubbard had malaria during [Feb3:Feb19]; this implies the propositions: (S_{1a}) "Jones asserts on Apr1 that Hubbard *had* malaria during exactly [Feb3:Feb19]." and (S_{1b}) "Jones asserts on May17 that Hubbard *did not have* malaria during exactly [Feb3:Feb19]." Similarly, $tp_2^{V,T}$ implies: (S_2) "Jones asserts on May17 that Hubbard *had* dengue during exactly [Feb3:Feb24]." We will see later that (S_{1a}), (S_{1b}) and (S_2) are ambiguous in several respects.

3. Bridge with referent-tracking systems

3.1. IUI repository and referent-tracking database

A referent tracking (RT) system is composed of two parts [1]. First, an IUI (Instance Unique Identifier) repository, which is an inventory of identifiers for individual entities, such as a specific patient Mr. Williams, his heart, his atrial fibrillation disease, and each of his atrial fibrillation episode. Second, the referent-tracking database ("RT-DB"), which is an inventory of assertions concerning the relationships between particulars, as well as between particulars and universals, and the ways those change over time. In the following, we write (as in [1]) "$\mathbf{IUI_{A}}$" for referring to the particular unique identifier $\mathbf{IUI_{A}}$, and "$\mathbf{\#IUI_{A}}$" for referring to the entity referred to by $\mathbf{IUI_{A}}$. For example, if $\mathbf{IUI_{Jones}}$ is the IUI referring to Jones, then $\mathbf{\#IUI_{Jones}}$ is Jones. In an ontological context, we will also write in bold the names of particulars and relations involving at least a particular, and use italic for universals.

3.2. Bridging B-RDB with RT-DB

An RT-DB can include a variety of kind of tuples [1,6], such as: PtoP tuples that each state a relation between particulars; PtoU tuples that each state an instantiation of a universal by a particular; and PtoLackU tuples that each state the lack of instantiation of a universal by a particular. Let's introduce $\mathbf{IUI_{disease\_H}}$ as referring to the disease that caused Hubbard's February symptoms of fever and nausea. (S_{1a}) (as defined above in section 2.2) can be expressed by a combination of two RT tuples:

- $\mathbf{PtoP_1}$ = < $\mathbf{IUI_{Jones}}$, **Apr1, inheres_in, RO,** ($\mathbf{IUI_{disease\_H}}$,$\mathbf{IUI_{Hubbard}}$), **[Feb3:Feb19]** >, which describes that Jones asserts on Apr1 that $\mathbf{\#IUI_{disease\_H}}$ **inheres_in Hubbard** during [Feb3:Feb19] (where **inheres_in** is a relation of RO, the Relation Ontology [7]).
- $\mathbf{PtoU_1}$ = < $\mathbf{IUI_{Jones}}$, **Apr1, inst, DO,** $\mathbf{IUI_{disease\_H}}$, *Malaria*, **[Feb3:Feb19]** >, which describes that Jones asserts on Apr1 that $\mathbf{\#IUI_{disease\_H}}$ **instance_of** *Malaria* during [Feb3:Feb19] (where *Malaria* is a class of DO, the Disease Ontology).

Altogether, those two tuples describe that Jones asserts on Apr1 that there is an instance of *Malaria* (namely, $\mathbf{\#IUI_{disease\_H}}$) inhering in Hubbard during [Feb3:Feb19]. We will now show ambiguities in the B-RDB tuples by describing the situation with RT-tuples.

4. The ambiguities of T-RDBs

4.1. Ambiguity 1: Having an asymptomatic disease vs. not having a disease

A first ambiguity is revealed when trying to express (S_{1b}) by RT tuples. (S_{1b}) can mean that Jones stated on May17 that $\mathbf{\#IUI_{disease\_H}}$ was not an instance of *Malaria*, in which case it is synonymous with the RT tuple $\mathbf{PtoLackU_1}$ = < $\mathbf{IUI_{Jones}}$, **May17, identical_with, RO,** $\mathbf{IUI_{disease\_H}}$, *Malaria*, **[Feb3:Feb19]** > (which describes that Jones asserted on May17 that there is no instance of *Malaria* that is **identical_with** $\mathbf{\#IUI_{disease\_H}}$). Alternatively, ($S_{1b}$) can mean that Jones stated on May17 that Hubbard did not have *any* instance of malaria during [Feb3:Feb19], in which case it is synonymous with the RT tuple $\mathbf{PtoLackU_2}$ = < $\mathbf{IUI_{Jones}}$, **May17, inheres_in, RO,** $\mathbf{IUI_{Hubbard}}$, *Malaria*, **[Feb3:Feb19]** >. $\mathbf{PtoLackU_2}$ asserts a stronger statement than the one asserted by $\mathbf{PtoLackU_1}$, as the former logically implies the latter. The difference between $\mathbf{PtoLackU_1}$ and $\mathbf{PtoLackU_2}$ is medically relevant: if Hubbard suffered during [Feb3:Feb19] from an asymptomatic malaria while having at the same time a dengue fever that caused his symptoms of fever and nausea, $\mathbf{PtoLackU_1}$ would hold, but not $\mathbf{PtoLackU_2}$; on the other hand, if Hubbard only had dengue fever during [Feb3:Feb19] and no malaria, $\mathbf{PtoLackU_2}$ would hold. But it is ambiguous whether (S_{1b}) means $\mathbf{PtoLackU_1}$ or $\mathbf{PtoLackU_2}$[2].

4.2. Ambiguity 2: Re-categorization of a formerly considered disease vs. consideration of a new disease

A second ambiguity is revealed when trying to express the B-RDB tuple $tp_2^{V,T}$=(Hubbard, Dengue, [Feb3:Feb24], Jones, [May17:ufn]) with RT tuples. In the case at hand, $\mathbf{IUI_{disease\_H}}$ had been defined as the IUI referring to Hubbard's disease that caused his symptoms of high fever and nausea in February[3]. Then $tp_2^{V,T}$ describes that Jones stated on May17 two things about $\mathbf{\#IUI_{disease\_H}}$:

- it inhered in Hubbard from Feb3 to Feb24, as expressed by $\mathbf{PtoP_2}$ **= < $\mathbf{IUI_{Jones}}$; May17 ; inheres_in ; RO ; ($\mathbf{IUI_{disease\_H}}$,$\mathbf{IUI_{Hubbard}}$) ; [Feb3:Feb24] >**
- it was an instance of *Dengue*, as expressed by $\mathbf{PtoU_2}$ **= < $\mathbf{IUI_{Jones}}$; May17 ; inst ; DO ; $\mathbf{IUI_{disease\_H}}$; *Dengue*; [Feb3:Feb24] >**

Suppose now that we are not in "scenario A", but in a "scenario B", that differs from scenario A in two respects. First, Jones learns on May17 that Hubbard's description of his February symptoms of high fever and nausea on April 1st were a lie (or a joke) – he never had them, and thus never had malaria nor dengue in February; second, he learns on May 17 that Hubbard had during [Feb3:Feb24] an unrelated hyperthyroidism. Instead of writing $tp_2^{V,T}$, Jones would introduce in the log of his B-RDB the tuple $tp_3^{V,T}$=(Hubbard, Hyperthyroidism, [Feb3:Feb24], Jones, [May17:ufn]).

In such a case, we should not use $\mathbf{IUI_{disease\_H}}$ to refer to Hubbard's hyperthyroidism, as the latter is unrelated to the fever and nausea symptoms Hubbard allegedly had in February. Instead, $\mathbf{IUI_{disease\_H}}$ would be an IUI without a reference[4]; and we would create a new IUI (let's say $\mathbf{IUI_{disease\_H\_2}}$) to refer to Hubbard's hyperthyroidism, such that Jones states on May17 two things about $\mathbf{\#IUI_{disease\_H\_2}}$:

- it inhered in Hubbard from Feb3 to Feb24, as expressed by $\mathbf{PtoP_3}$ **= < $\mathbf{IUI_{Jones}}$; May17 ; inheres_in ; RO ; ($\mathbf{IUI_{disease\_H\_2}}$,$\mathbf{IUI_{Hubbard}}$) ; [Feb3:Feb24] >**
- it was an instance of *Hyperthyroidism*, as expressed by $\mathbf{PtoU_3}$ **= < $\mathbf{IUI_{Jones}}$; May17 ; inst ; DO ; $\mathbf{IUI_{disease\_H\_2}}$; *Hyperthyroidism* ; [Feb3:Feb24] >**

Note that the B-RDB descriptions of scenarios A and B are similar (the only difference between $tp_3^{V,T}$ and $tp_2^{V,T}$ is the use of "*Hyperthyroidism*" instead of "*Dengue*"); but this does not describe an important difference between scenarios A and B, revealed by the RT descriptions of both scenarios: the RT description of scenario A uses the formerly introduced $\mathbf{IUI_{disease\_H}}$ (with $\mathbf{PtoP_2}$ and $\mathbf{PtoU_2}$), whereas the RT description of scenario B uses the newly introduced $\mathbf{IUI_{disease\_H\_2}}$ (with $\mathbf{PtoP_3}$ and $\mathbf{PtoU_3}$).

---

[2] Note that in both cases, an additional RT-tuple should be introduced, stating that the universal-tuple $\mathbf{PtoU_1}$ above was wrong – see [8] for a suggestion of how to represent this as a D-tuple.

[3] This example raises a more general issue, namely how we define the identity of a particular disease. In the case at hand, the particular $\mathbf{\#IUI_{disease\_H}}$ is defined as the disease causing some symptoms of Hubbard (high fever, nausea) during an approximate time period (during the month of February).

[4] This mistake is named "(A1)" in [8].

4.3. Ambiguity 3: One vs. several pathological processes

A third ambiguity concerns the number of entities. Consider the following relation $R^{PP,V}$ (with "DIS” replaced by "PP”, referring to “pathological process”) :

PAT	*PP*	*@V*
Williams	AF episode	[Jun6:Jun15]

Note first that the proposition associated to this tuple cannot be ‘Williams had an AF episode that spanned exactly the time interval [Jun6:Jun15]’, as this describes a so-called “telic” fact that cannot be represented in point-based semantics [9]. The semantic of such predicate would rather be ‘Williams had one or several AF episodes that together spanned a time interval including [Jun6:Jun15]’. Therefore, Williams may have had only one AF episode spanning the whole interval [Jun6:Jun15]. Alternatively, he may have had two AF episodes: a first one during [Jun6:Jun10], and a second one during [Jun11:Jun15]. There is thus an ambiguity concerning how many AF episodes Williams had. Note that such ambiguity would not appear in a RT system: each AF episode would have its own IUI, therefore the RT data would explicitly describe if Williams had one or several AF episode(s) during the interval [Jun6:Jun15].

5. Conclusion

This article showed several ambiguities concerning the individual entities underlying the predicates expressed by T-RDBs when the values of some of their attributes (such as *DIS* in R^V or *PP* in $R^{PP,V}$) refer to classes or universals, rather than to individuals; this makes it a particular case of a difficulty named the “assumption of inherent classification” [11, 7]. This suggests that to avoid such ambiguities, medical T-RDBs could be structured on the basis of ontology-based representations that deal carefully with the particulars involved, such as referent tracking systems.

References

[1] W. Ceusters, B. Smith, Strategies for referent tracking in electronic health records, *Journal of Biomedical Informatics* **39** (2006), 362–378.

[2] C. Combi, E.T. Keravnou, Y. Shahar, *Temporal Information Systems in Medicine*, Springer, 2010.

[3] C.J. Date, H. Darwen, N.A. Lorentzos, *Time and relational theory: temporal databases in the relational model and SQL*, Morgan Kaufmann, Waltham, MA, 2014.

[4] B. Smith, W. Ceusters, Ontological realism: A methodology for coordinated evolution of scientific ontologies, *Applied Ontology* **5** (2010), 139–188.

[5] A. Varzi, Mereology, in: E.N. Zalta (Ed.), *Stanf. Encycl. Philos.*, Winter 2016, Metaphysics Research Lab, Stanford University, 2016.

[6] W. Ceusters, P. Elkin, B. Smith, Negative findings in electronic health records and biomedical ontologies: a realist approach, *International Journal of Medical Informatics* **76** (2007), S326–S333.

[7] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, et al., Relations in biomedical ontologies, *Genome Biology* **6** (2005), R46.

[8] W. Ceusters, Dealing with mistakes in a referent tracking system, *Proceedings of Ontology for the Intelligence Community* (2007), 5–8.

[9] P. Terenziani, R.T. Snodgrass, Reconciling point-based and interval-based semantics in temporal relational databases: A treatment of the telic/atelic distinction, *IEEE Transactions on Knowledge and Data Engineering* **16** (2004), 540–551.

[10] J. Parsons, Y. Wand, Emancipating instances from the tyranny of classes in information modeling, *ACM Transactions on Database Systems (TODS)* **25** (2000), 228–268.