Reasoning with enhanced Temporal Entity-Relationship Models (Extended Abstract)

Alessandro Artale
Department of Computation
UMIST, Manchester, UK
artale@irst.itc.it

Enrico Franconi
Department of Computer Science
University of Manchester, UK
franconi@cs.man.ac.uk
http://www.cs.man.ac.uk/~franconi/

Abstract

This preliminary work gives a logical formalisation of the various properties that characterise and extend different temporal ER models which are found in literature. The formal language we propose is a member of the family of Description Logics (DL), which have been proved useful for a logical reconstruction of the most popular conceptual data modelling formalisms, including the (enhanced) ER model. The proposed DL has the ability to express both enhanced temporal ER schemas and Integrity Constraints (IC) in the form of general axioms imposed on the schema itself. The devised logic has a decidable reasoning problem, thus allowing for automated reasoning over the whole conceptual representation, which includes both the ER schema and the ICs over it.

1 Introduction

Temporal Databases are databases that store historical informations, i.e., essentially past and present data. In the temporal-ER community two different modelling approaches exist to providing temporal support. The *implicit* approach hides the temporal dimension in the interpretation structure for the ER constructs. Essentially, the ER constructs, with no distinction between the temporal and the atemporal ones, are always interpreted with a temporal semantics, so that, for example, instances of temporal relationships are always potentially time-varying tuples. More complex and challenging becomes the case where an *explicit* approach to providing temporal support is adopted. The explicit approach retains the original semantics of ER constructs while adding new constructs that allow the modeller to introduce explicit timestamps to indicate temporal entities and temporal relationships, and to constrain their temporal properties.

If the implicit approach has the attractive of leaving unchanged the original ER model, "this approach rules out the possibility of designing non-temporal databases or databases where some part of a database is non-temporal and the rest is temporal" [Gregersen and Jensen, 1999].

The formalism introduced in this paper follows an implicit approach, but extends it by providing the possibility to explicitly express temporal constructs. The formalisms is centred around a basic temporal ER model, which does not have any new syntactical constructs with respect to a "standard" ER model. In addition, it is possible to express additional explicit temporal statements in the form of expressive *integrity constraints* to the model, and in particular it is possible to state as an integrity constraint the distinction between time-varying and snapshot (i.e., time invariant) constructs. This last requirement – called *upward compatibility* – is crucial if we want to maintain unchanged the meanings of conventional (legacy) ER diagrams also when used inside a temporal model.

2 The Description Logic

The formal language used in this work is an expressive temporal Description Logic. Advantages of using Description Logics are their high expressivity combined with desirable computational properties – such as decidability, soundness and completeness of deduction procedures. The core is the Description Logic $\mathcal{ALCQI}$, which is able to capture conventional ER models and it has the ability to express a powerful class of integrity constraints. The core language is then extended with a tense logic, thus allowing to express also temporal integrity constraints ($\mathcal{ALCQIT}$). Reasoning in the whole framework is a *decidable* task, allowing for a complete calculus for temporal integrity constraints. In this extended abstract we will not describe the temporal description logic; the final version of the paper will introduce the language in a formal way.

$\mathcal{ALCQI}$, the non-temporal part of the language introduced in this paper, was designed such that it is able to encode database schemas expressed in the most interesting Semantic Data Models and Object-Oriented Data Models [Calvanese *et al.*, 1998; Franconi and Sattler, 1999]. Recently, strictly more expressive conceptual data models based on DLs have been considered, most notably the $\mathcal{DLR}$ conceptual data modelling formalism. $\mathcal{DLR}$ was first introduced by [De Giacomo and Lenzerini, 1999] as a means for encoding expressive semantic data models for information systems such as extended Entity Relationship in the context of schema integration. We have chosen to limit the expressivity of $\mathcal{DLR}$ since we are looking for a language implementable with the current technology, but still capable to encode an interesting enhancement of the ER formalism. In particular, we have experimented it using the current academic implementations of expressive DLs, namely the systems FaCT [Horrocks, 1998] and iFaCT (FaCT extended with inverse relations). It has been recently demonstrated [Horrocks and Patel-Schneider, 1999] that the logic we are considering here allows for the implementation of sound and complete reasoning algorithms that behave quite well both in realistic applications and systematic tests.

3 Temporal Entity-Relationship Models

We refer to an extended version of the basic temporal ER model, which includes taxonomic relationships, arbitrary boolean constructs, entity definitions by means of either necessary or sufficient conditions or both, and (temporal) integrity constraints expressed by means of generalised axioms. We have shown how a schema expressed in this conceptual data model can be expressed in a $\mathcal{ALCQIT}$ knowledge base – whose models correspond with legal database states of the ER diagram – allowing for reasoning services such as satisfiability of a schema or the computation of a logically implied integrity

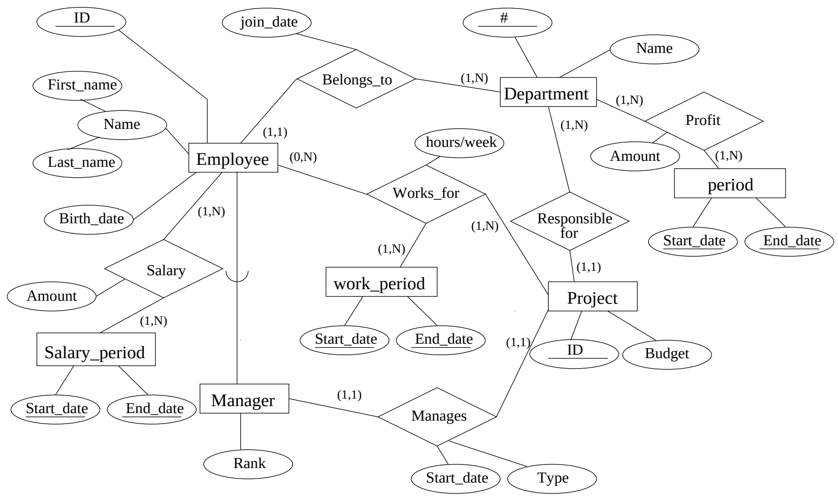

Figure 1: The reference temporal ER diagram

constraint. It is important to emphasise the fact that in our approach the integrity constraints are part of the schema, and that reasoning is carried on by taking in complete account all the information contained in the schema.

Based on the results of [Calvanese *et al.*, 1994; 1998], we can prove that the translation from ER to the DL is correct, in the sense that whenever a reasoning problem has a specific solution in the ER model, then the corresponding reasoning problem in the DL has a corresponding solution, and vice-versa. This is grounded on the fact that there is a precise correspondence between legal databases of $\mathcal{D}$ (intuitively, a legal database is a database – i.e., a finite relational structure – which conforms to the constraints imposed by the schema) and models of Σ. Thus, it is possible to exploit reasoning procedures in the DL for solving reasoning problems in the ER model. The reasoning problems we are mostly interested in are *consistency* of a ER schema – which is mapped to a satisfiability problem in the corresponding DL knowledge base – and *logical implication* within a ER schema – which is mapped to a logical implication problem in the corresponding DL knowledge base.

Let us consider the example ER diagram of figure 1; this diagram is the running example considered in the survey paper [Gregersen and Jensen, 1999]. We first translate (a fragment of) the ER schema in the description logic knowledge base Σ_{ER}:

$\texttt{WORKS-FOR} \sqsubseteq \texttt{has-prj} : \texttt{Project} \ \& \ \texttt{has-emp} : \texttt{Employee}$

$\texttt{Project} \sqsubseteq \exists \texttt{has-prj}^{-1}.\texttt{WORKS-FOR}$

$\texttt{Manager} \sqsubseteq \texttt{Employee}$

We now want to impose additional integrity constraints, which are expressed by means of terminological axioms in a knowledge base Σ_{IC}:

- Managers are the only employees who do not work for a projects (she/he just manages it):

  $\texttt{Employee} \ \& \ \forall \texttt{has-emp}^{-1}.\neg\texttt{WORKS-FOR} \sqsubseteq \texttt{Manager}$

- A manager becomes qualified after a period when she/he was just an employee:

  $\texttt{Manager} \sqsubseteq \texttt{Qualified} \ \mathcal{S} \ (\texttt{Employee} \ \& \ \neg\texttt{Manager})$

It turns out that the following integrity constraints are logically implied from $\Sigma_{ER} \cup \Sigma_{IC}$:

- For every project, there is at least an employee who is not a manager:

  $\texttt{Project} \sqsubseteq \exists(\texttt{has-prj}^{-1} \circ \texttt{has-emp}).\neg\texttt{Manager}$

- A manager worked in a project before managing some (possibly different) project:

  $\texttt{Manager} \sqsubseteq \Diamond^{-}\exists(\texttt{has-emp}^{-1} \circ \texttt{has-prj}).\texttt{Project}$

Please note that these deductions are not trivial, since from the ER schema the cardinality constraints do not impose that employees necessarily work in a project. A reasoner based on $\mathcal{ALCQIT}$ can automatically derive the validity of the above integrity constraints.

4 Upward Compatible Models

In the full paper it will be sketched how the presented temporal DL could formalise upward-compatible temporal ER models by simply imposing additional specific integrity constraints on the interrelations between snapshots and temporal constructs. In particular, we suppose that both entities and relationships can be *snapshot* marked (i.e., each of their instances have a global lifetime, as in the case they derive from a legacy diagram); *temporary* marked (i.e., each of their instances has a temporary lifetime) and *unmarked* (i.e., the user doesn't want to stress the temporal validity). Note that, with this marking approach, legacy diagrams should be snapshot marked before inserting them in a temporal model to maintain the upward compatibility.

In the temporal database community, a relation – in the temporal relational model – is called *heterogeneous* when its attribute values have different period of existence [Tansel and Tin, 1998]. Both temporary and unmarked relationships admit heterogeneous values. However, they can be forced to have *homogeneous* tuples [Gadia, 1988] – i.e., the values of a tuple are valid on the same time period – by imposing additional specific integrity constraints. It should be noted that an extension of the Description Logic with *global roles* is needed.

[McBrien *et al.*, 1992] enforce temporal constraints on the validity time of instances of entities involved in heterogeneous relationships. The original ERT model [Theodoulidis *et al.*, 1991] has been extended to include *historical* marks (H-mark) by imposing particular temporal constraints on the validity time of the different ER-roles. Different temporal constraints give rise to different H-marked relationships. For example, a relationship between entities E_1 and E_2 is H-marked *past* if the instances of E_1 involved in the relationship hold at intervals before the intervals where the instances of E_2 hold. To represent H-marked relationships a move from a point-based to an interval-based temporal DL is necessary. Interval temporal DLs have been proposed in literature [Artale and Franconi, 1998; Haarslev *et al.*, 1998]. While they could be the basis for a formal framework to enforce integrity constraints on historical relationships, we are currently analysing the tradeoff between their expressive power and decidability of deduction (see [Artale and Franconi, 1999] for more details).

5 Conclusions

This preliminary work gives a logical formalisation of a temporal ER model, which has the ability to express both enhanced temporal ER schemas and (temporal) integrity constraints in the form of general axioms imposed on the schema itself. The formal language we have proposed is a member of the family of Description Logics, and it has a decidable reasoning problem, thus allowing for automated deduction over the whole conceptual representation.

We have also shown how the integrity contraints can encode the distinction between time-varying and snapshot constructs.

This work is just at the beginning. The most promising research direction to be explored is to better characterize the expressivity of temporal integrity constraints in order to axiomatise several extensions as proposed in the literature of temporal ER models.

Acknowledgements

The work presented in this paper was partially supported by the "Foundations of Data Warehouse Quality" (*DWQ*) European ESPRIT IV Long Term Research (LTR) Project 22469.

References

[Artale and Franconi, 1998] Alessandro Artale and Enrico Franconi. A temporal description logic for reasoning about actions and plans. *Journal of Artificial Intelligence Research*, 9:463–506, 1998.

[Artale and Franconi, 1999] Alessandro Artale and Enrico Franconi. Temporal description logics. In L. Vila, P. vanBeek, M. Boddy, M. Fisher, D. Gabbay, A. Galton, and R. Morris, editors, *Handbook of Time and Temporal Reasoning in Artificial Intelligence*. MIT Press, 1999.

[Calvanese *et al.*, 1994] Diego Calvanese, Maurizio Lenzerini, and Daniele Nardi. A unified framework for class-based representation formalisms. In *Proc. of the 4 th International Conference on Principles of Knowledge Representation and Reasoning*, Bonn, Germany, May 1994.

[Calvanese *et al.*, 1998] D. Calvanese, M. Lenzerini, and D. Nardi. Description logics for conceptual data modeling. In Jan Chomicki and Günter Saake, editors, *Logics for Databases and Information Systems*. Kluwer, 1998.

[De Giacomo and Lenzerini, 1999] G. De Giacomo and M. Lenzerini. Description logic framework for information integration. In *Proceedings of the 6th International Conference on the Principles of Knowledge Representation and Reasoning (KR-98)*, pages 2–13. Morgan Kaufmann, 1999.

[Franconi and Sattler, 1999] Enrico Franconi and Ulrike Sattler. A data warehouse conceptual data model for multidimensional aggregation: a preliminary report. *Italian Association for Artificial Intelligence AI*IA journal*, 1999. To appear.

[Gadia, 1988] S.K. Gadia. A homogeneous relational model and query languages for temporal databases. *ACM Transactions On Database Systems*, 13:418–448, 1988.

[Gregersen and Jensen, 1999] H Gregersen and J. S. Jensen. Temporal Entity-Relationship models - a survey. *IEEE Transactions on Knowledge and Data Engineering*, 1999. To appear.

[Haarslev *et al.*, 1998] Volker Haarslev, Carsten Lutz, and Ralf Möller. Foundation of spatiotemporal reasoning with description logics. In A.Cohn, L.Schubert, and S.C.Shapiro, editors, *Proc. of the 6 th International Conference on Principles of Knowledge Representation and Reasoning*, pages 112–123, Trento, Italy, June 1998. Morgan Kaufmann.

[Horrocks and Patel-Schneider, 1999] I. Horrocks and P. F. Patel-Schneider. Optimising description logic subsumption. *Journal of Logic and Computation*, 1999. To appear.

[Horrocks, 1998] I. Horrocks. Using an expressive description logic: FaCT or fiction? In *Proc. of the 6 th International Conference on Principles of Knowledge Representation and Reasoning*, pages 636–647, Trento, Italy, 1998.

[McBrien *et al.*, 1992] P. McBrien, A.H. Seltveit, and B. Wangler. An Entity-Relationship model extended to describe historical information. In *Proc. of CISMOD'92*, pages 244–260, Bangalore, India, 1992.

[Tansel and Tin, 1998] A.U. Tansel and E. Tin. Expressive power of temporal relational query languages and temporal completeness. In O. Etzion, S. Jajodia, and S. Sripada, editors, *Temporal Databases - Research and Practice*, Lecture Notes in Computer Science, pages 129–149. Springer-Verlag, 1998.

[Theodoulidis *et al.*, 1991] C. Theodoulidis, P. Loucopoulos, and B. Wangler. A conceptual modelling formalism for temporal database applications. *Information Systems*, 16(3):401–416, 1991.