Department of
Information Engineering
and Computer Science
DISI

UNIVERSITY
OF TRENTO - Italy

DISI - Via Sommarive 5 - 38123 Povo - Trento (Italy)
http://disi.unitn.it

FROM ER MODELS TO THE ENTITY MODEL

Fausto Giunchiglia and Mattia Fumagalli

Date (2014-October)

Technical Report # DISI-14-014

From ER Models to the Entity Model

Fausto Giunchiglia, Mattia Fumagalli

Department of Information Engineering and Computer Science,
University of Trento, via Sommarive 14 I-38123, Povo, Trento, Italy
{fausto.giunchiglia,mattia.fumagalli}@unitn.it

Abstract. In this paper, a new knowledge representation formalism, called the *entity model*, is introduced. This model can be used to address knowledge diversity by making the modeling assumptions of different knowledge representations explicit and by rooting them in a world representation. The entity model can be used to: 1) detect the possible ways in which the diversity appears in ER models and therefore improving their representational adequacy; 2) make the modeling assumptions behind different ER models explicit; 3) combine the different ER models in a unified view, thus enabling data integration.

Keywords: knowledge representation, conceptual modeling, ontological analysis, semantic heterogeneity, knowledge diversity.

1 Introduction

When we set ourselves for representing the world we have to deal with what we often informally call *diversity*. On one side, many criteria can be found for distinguishing categories of things which can make two representations of the same portion of reality completely incompatible. On the other side, diversity allows us to identify the single items which are needed in order to distinguish the different portions of reality. In the first case, diversity can be conceived as an intrinsic property of our knowledge representation, namely a function of local factors, like needs, beliefs or culture [4]. In the second case, diversity can be conceived as a property of things in the world, namely the characteristic of being distinct.

Diversity in world representations can be reduced to what is often called *semantic heterogeneity*. Semantic heterogeneity is a long-standing problem [5], for which a comprehensive solution still does not exist, that needs to be addressed in different application areas such as resource discovery, data integration, data migration, query translation, peer-to-peer networks, agent communication, schema and ontology matching. For this problem we propose to study semantic heterogeneity as the result of the projection of diversity in the real world into further, possibly diverse, representational choices. In other terms, we want to study the semantic heterogeneity problem as the result of injecting possibly diverse representational choices into the pre-existing real world diversity. To this extent, we propose a new formalism, and a corresponding methodology, called the *entity model*, which represents the real world diversity, and we take the ER model [2] as the formalism used to encode diversity in world representations.

Our work can also be seen as providing an ontological foundation to the ER model. So far, there have been very few attempts at using ontological distinctions for constraining the semantics of a conceptual modelling language. The most complete research in this direction has been made in [6], where an ontological foundation for UML language is proposed. The ER model still needs to be ontologically well-founded. Moreover, as far as we know, there are no attempts to constrain a conceptual modeling language by means of a reference ontology which takes into account the distinction between world diversity and knowledge diversity.

The rest of the paper is structured as follows. Section 2 provides the motivation for our work showing an example of diversity in ER modeling and introducing the relations between the ER models and the entity model. Section 3 gives an explanation of the basic steps for creating the entity model and for rooting different ER models in the entity model.

2 Motivation

One concrete example of how diversity can be conceived as a structural feature of ER modelling is described through Fig. 1.

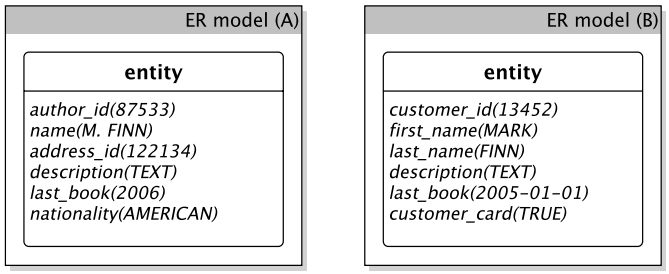

Fig. 1. Different representations for the same portion of reality

The matching of two references extracted from two different datasets (see Fig. 1) raises fundamental ontological and meta-ontological issues, for instance: (i) what entities do we have? (ii) Are these the same entities? (iii) What are the admissible attributes for these entities?

The entity model, taking into account the distinction between diversity in knowledge representation and the real world diversity, provides a model of the world based on the individuation of entities. The key advantage of this representation formalism is that, by rooting diversity of representations in the diversity of the world, we can trace back all the possible sources of diversity and therefore select the possible ways this diversity appears in ER models (Fig. 2). By tracing diversity, it is possible to verify and evaluate the potential of the entity model, which can be used for: improving the ER models representational adequacy;

making the modelling assumptions behind different ER models explicit; combining different ER models in a unified view.

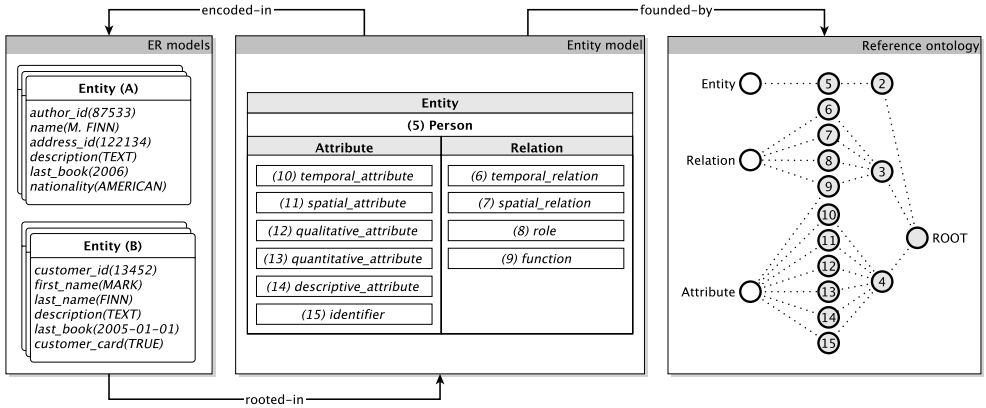

Fig. 2. Rooting the diversity of representations in the diversity of the world

The entity model has a straightforward formalization into the ER model. The set of terms *T* denoting the entity model basic components can be seen as a triple, *T* = <*E, R, A*> [3]. Each term of *T* can be conceived as belonging to a hierarchy of terms rooted in a reference ontology, based on the theory of knowledge provided by Aristotle and Brentano [1] and used to ground the entity model. The entity model defines the set of *core entities, core attributes* and *core relations*, which are derived from the reference ontology. Such core components will constitute the structure to be used for rooting the ER models in the entity model and therefore for tracing diversity. For instance, let us consider the example of Fig. 2. Entities (A) and (B) are derived from the composition of terms denoting entities, relations and attributes. All such terms can be traced back to a core component of the entity model: terms like CUSTOMER and AUTHOR can be traced back to ROLE; terms like LAST_BOOK can be traced back to TEMPORAL_ATTRIBUTE, which may be specified as DATE_OF_PURCHEASE or DATE_OF_CREATION. The entity model will provide a set of *basic grounding constraints* for guiding the modelling choices of the knowledge engineer. For instance, let us assume that according to the entity model all the *entity terms* are those terms α, such that *'x is* α*' iff* α *is needed to be x*. A term like CUSTOMER should not be considered as an entity term (e.g., *Mark,* may be a CUSTOMER, but it is not needed to be a CUSTOMER for being *Mark*). Moreover, the ER models may provide information for extending the entity model and its reference ontology. For instance, AUTHOR_ID and CUSTOMER_ID may be used to extend the class IDENTIFIER. Consequently, through the entity model it is possible to capture how (A) and (B) reflect different representational choices for describing the same individual and to combine these different representations in a unified view.

3 Our approach

Our work is based on a methodology, to be divided in five main steps, which encompasses: (1) how to ground the entity model on a philosophical theory; (2) how to de-sign the reference ontology, starting from (1) with a typical top-down process; (3) how to create the entity model (encoding it in the ER model); (4) how to build ER models, with a bottom-up approach, from multiple sources of data, which are compatible with the entity model; (5) how to feedback the lessons learned in the bottom-up approach into possible modifications of the ER models and possibly even the entity model, thus extending the reference ontology.

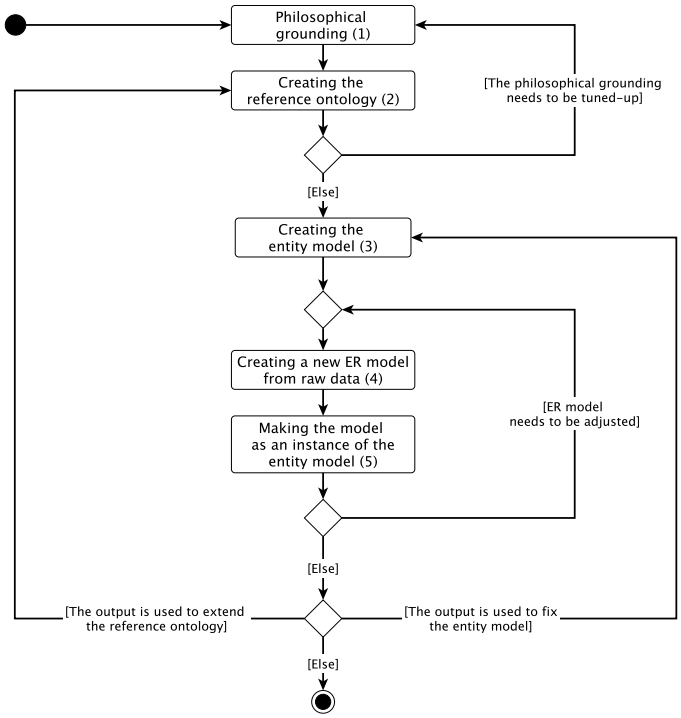

Fig. 3. From ER models to the entity model

All the steps (Fig. 3) of the methodology can be grouped into two main processes: a *global process*, or a set of top-down steps, covering (1), (2) and (3), whose output is an encoding of diversity in the real world, and a *local process*, namely a set of bottom-up steps, covering (4) and (5), whose output is an encoding of knowledge diversity, namely an expandable set of ER models. The entity model is generated, updated and extended through both these processes.

4 Acknowledgements

This research has received funding from the Smart Society Collaborative Project, funded by the European Commission's 7th Framework ICT Programme for Research and Technological Development under the Grant Agreement no. 600854.

References

1. Brentano, F. On the Several Senses of Being in Aristotle. UC Press (1976)
2. Chen, P. The Entity-Relationship Model - Toward a Unified View of Data. In: ACM Transactions on Database Systems, pp. 9-36 (1976)
3. Giunchiglia, F., Dutta, B., Maltese, V. From Knowledge Organization to Knowledge Representation. In: ISKO UK (2013)
4. Giunchiglia, F., Maltese, V., Dutta, B. Domains and context: First steps towards managing diversity in knowledge. In: Journal of Web Semantics: Science, Services and Agents on the World Wide Web, Volume 12-13, pp. 53-63 (2012)
5. Guarino, N. The Ontological Level: Revisiting 30 Years of Knowledge Representation. In: Conceptual Modeling: Foundations and Applications, pp. 52-67 (2009)
6. Guizzardi, G. Ontological Foundations for Structural Conceptual Models. In: Telematica Instituut Fundamental Research Series (2005)