

Conceptual Modeling: the Linguistic Approach

Lucia Castro¹

Fernanda Araujo Baião¹

Giancarlo Guizzardi²

Abstract: After more than thirty years of its first introduction, conceptual modeling remains an important research field, which has been recently addressed by the literature on semantic interoperability in its various forms (model integration, service interoperability, knowledge harmonization, taxonomy alignment), domain engineering and the creation of conceptual models through Natural Language Processing (NLP), to name a few. In the database conceptual design, the designer must learn the language used in the Universe of Discourse (UoD) to be modeled, along with its underlying concepts, and then represent such concepts in a modeling language. Thus, the conceptual modeling process can be seen as a translation. For the resulting model to be both detailed and unambiguous, the designer must represent the UoD in a generative language which constructs can convey the same concepts represented in the respective natural language. For the whole process to be effective, we argue that the adoption of modeling languages and methodologies that are based on well-founded ontological theories is required. We propose the use of a linguistic approach for conceptual modeling from natural language texts, and illustrate how it may be applied using the well-founded modeling language *OntoUML*.

1 Conceptual Modeling

Batini et al. [1] state that the conceptual design “*starts from the specification of requirements and results in the conceptual schema of the database*” and “*is by far the most critical phase of database design and further development of database technology is not likely to change this situation*”, that it is an activity that *cannot be very much helped by automatic tools.*” that “*the designer has full responsibility for the process of understanding and transforming requirements into conceptual schemas*”. According to these statements, the conceptual modeling process is a subjective, time consuming activity that relies on the experience of a professional rather than in a prescriptive methodology.

¹ NP2Tec/UNIRIO, Rio de Janeiro, Brazil

{lucia.castro, fernanda.baiao}@uniriotec.br

² NEMO/UFES, Espírito Santo, Brazil

{gguizzardi@inf.ufes.br}

In order to make conceptual modeling a more objective process, Chen [2] enunciated eleven rules for the translation of data requirements from a natural language (in this case English) to an ER model. Based on such rules, many projects have proposed tools and heuristics to automate the modeling activity. Examples can be found in [3] and [4].

2 The Linguistic Approach

The projects mentioned above work from the perspective of the meta-model structure; they turn to linguistics to try and establish rules as to what natural language construct should be mapped to which representation language construct. Although linguistics has a supportive role, such works tend to overlook semantic and syntactic details that are present in the natural language data requirement representation but that cannot be equally represented in the conceptual model. As so, representations tend to not be well evaluated in terms of correctness and completeness.

The first mapping rule presented by Chen [2], states, for instance, that “*A common noun in English corresponds to an entity type in an ER diagram*”. Although one can relate to this rule in a very natural way (after all, “employee” and “vehicle”, for instance, are indeed common nouns), words like “date”, “water”, “red” and “size”, to name a few, are also common nouns – should they all be represented as entities? From a linguistic perspective such subtleties become more evident, as does the need for a well-founded ontological language to represent them. From the linguistic perspective, accurate and complete conceptual data models must be represented in an ontological, generative language, in such a way that the reading of the model can be the same as that of the natural language represented requirements.

Using the same words of the example above, a well-founded ontological language as *OntoUML* [5] differentiates “employee” from “water”, for instance. The first common noun represents a kind, i.e., the construct that corresponds to concrete common nouns, a concept composed and identified by its characteristics. The second would be represented by a quantity, that inheres the idea of a mass (or uncountable) noun.

References

1. Batini, C., Ceri, S., Navathe, S., 1992. *Conceptual Database Design*, Benjamin/Cummings.
2. Chen, P., 1983. *English Sentence Structure and Entity-Relationship Diagrams*, Elsevier Science Publishing
3. Overmyer, S., Lavoie, B., Rambow, O., 2001. *Conceptual Modeling through Linguistic Analysis Using LIDA*, 23rd International Conference on Software Engineering.
4. Hartmann, S., Link, S., 2007. *English Structures and EER Modeling*, Fourth Asia-Pacific Conference on Conceptual Modeling.
5. Guizzardi, G., 2005. *Ontological Foundations for Structural Conceptual Models*, Telematica Instituut Fundamental Research Series 15, Universal Press.