Chapter XIX
Engineering Conceptual Data Models from Domain Ontologies: A Critical Evaluation

Haya El-Ghalayini
University of the West of England (UWE), UK

Mohammed Odeh
University of the West of England (UWE), UK

Richard McClatchey
University of the West of England (UWE), UK

ABSTRACT

This paper studies the differences and similarities between domain ontologies and conceptual data models and the role that ontologies can play in establishing conceptual data models during the process of information systems development. A mapping algorithm has been proposed and embedded in a special purpose Transformation Engine to generate a conceptual data model from a given domain ontology. Both quantitative and qualitative methods have been adopted to critically evaluate this new approach. In addition, this paper focuses on evaluating the quality of the generated conceptual data model elements using Bunge-Wand-Weber and OntoClean ontologies. The results of this evaluation indicate that the generated conceptual data model provides a high degree of accuracy in identifying the substantial domain entities along with their relationships being derived from the consensual semantics of domain knowledge. The results are encouraging and support the potential role that this approach can take part in the process of information system development.
INTRODUCTION

In the last decade, ontologies have been considered as essential components in most knowledge-based application development. As these models are increasingly becoming common, their applicability has ranged from the artificial intelligence domain, such as knowledge engineering/representation and natural language processing, to different fields like information integration and retrieval systems, the semantic Web, and the requirements analysis phase of the software development process. Therefore, the importance of using ontologies in building conceptual data models (CDMs) (CDMs) has already been recognized by different researchers. In our approach, we claim that the differences and similarities between ontologies and CDMs play an important role in the development of CDMs during the information system development process. We indicate that CDMs can be enriched by modeling the consensual knowledge of a certain domain, which, in turn, minimizes the semantic heterogeneities between the different data models (El-Ghalayini, Odeh, McClatchey, & Solomonides, 2005). We chose to study ontologies represented by the Web ontology language (OWL), since it is the most recent Web ontology language released by the World Wide Web Consortium in February 2004 (W3C-World Wide Web Consortium, 2005), and its formal semantics are based on description logics (DL).

The remainder of this article is structured as follows. The next section provides relevant information related to ontologies, CDMs, and the so-called transformation engine (TE). Then the following section discusses the process of evaluating the TE and its parameters, in general, and the qualitative dimension in evaluating the quality of the generated CDM elements, using ontological rules. This evaluation is demonstrated by a real-life case study related to the transparent access to multiple bioinformatics information sources (TAMBIS) ontology; finally the conclusion and future work are presented.

ONTOiOLOGY VS. CONCEPTUAL DATA MODEL

This section informally explores ontologies and CDMs, including their similarities and differences. The literature shows many definitions of ontologies with the most popular definition proposed by Gruber (1995) as “a formal, explicit specification of a shared conceptualization” (p. 907). In general terms, an ontology may be defined as expressing knowledge in a machine-readable form to permit a common understanding of domain knowledge, so knowledge can be exchanged between heterogeneous environments.

On the other hand, conceptual data models capture the meaning of information for modelling an application and offer means for organizing information for the purposes of understanding and communication (Mylopoulos, 1998). The major role of the CDM is to model the so-called universe of discourse (UoD), entities and relationships in relation to particular user requirements independent of implementation issues. Hirschheim, Klein, and Lyytinen (1995) define the Universe of Discourse in the information systems (IS) world as “the slice of reality to be modelled” (p. 58). Therefore, there are some similarities and differences between ontologies and CDMs. Both are represented by a modeling grammar with similar constructs, such as classes in ontologies that correspond to entity types in CDMs. Thus, the methodologies of developing both models have common activities (Fonseca & Martin, 2005). While ontologies and CDMs share common features, they have some differences. According to Guarino’s (1998) proposal of ontology-driven information systems, an ontology can be used at the development or run time of IS, whereas a CDM is a building block of the analysis and design process of an IS.

Moreover Fonseca and Martin (2005) define two criteria that differentiate ontologies from CDMs; the first is the objectives of modeling and the second is objects to model. Using the first criterion, an ontology focuses on the descrip-
