Using Ontology and Rule-Based Reasoning for Conceptual Data Models Synonyms Detection:
A Case Study

Ljubica Kazi, University of Novi Sad, Technical Faculty "Mihajlo Pupin," Zrenjanin, Serbia

Zoltan Kazi, University of Novi Sad, Technical Faculty "Mihajlo Pupin," Zrenjanin, Serbia

ABSTRACT

Conceptual data models can change during the information system development and teamwork phases, which require constantly monitoring with synonyms detection. This study elaborates on an approach for detecting synonyms in an entity-relationship model based on mapping with ontological elements. The use of a specific data model validator (DMV) tool enables formalization of the ontology and ER models, as well as their integration with the set of reasoning rules. The reasoning rules enable mapping between formalized elements of the ontology and ER model, and the extraction of synonyms. Formalized elements and reasoning rules are processed within Prolog for the extraction of synonyms. An empirical study conducted by using university student exams demonstrates usability of the proposed approach. The results show effectiveness in extraction of synonyms in all types of conceptual data model elements.

KEYWORDS

Automated Reasoning System, Data Consistency, ER Model, Extraction, Mapping Elements, Modeling Grammars, OWL, Reasoning Rules, Software Change Management, Structural Similarity

INTRODUCTION

Human interpretation in mapping real world concepts to abstraction instances in conceptual modeling grammars (Chua, Storey, & Chiang, 2012) leads to the problem of identifying relevant model elements (Athenikos & Song, 2013) and diversity of created conceptual data models (CDMs) for the same business domain (Jabbari, Sabegh & Recker, 2017), with a wide occurrence of synonyms. Therefore, conceptual data modeling is considered a critical phase of information system development (Hannola & Ovaska, 2011). In practice, conceptual data models are transformed into other implementation-oriented data models—relational schemas (An, Hu, & Song, 2010), and object-oriented data models, i.e. UML class diagrams (Keet, & Fillottrani, 2015). Synonyms in conceptual data models should be managed carefully, because of their possible transfer to database and software elements.

This paper elaborates on the approach to synonyms detection in a conceptual model by using ontology and automated reasoning, which is briefly introduced in Kazi, Kazi, and Radulovic (2016). The approach is based on mapping elements of the conceptual data model (CDM) created with ER modeling grammar to elements of ontology, both created within the same domain. Both models are created using appropriate tools (ER model within CASE tool and ontology within ontology tool). A specially created Data Model Validator (DMV) tool transforms their original form (XML form of CDM model and OWL form of ontology) into Prolog-like clauses and integrates them with the externally stored reasoning rules written it Prolog-suitable form. This way, integrated Prolog program is created within the DMV tool. This tool passes the Prolog program into Prolog for processing queries, which extracts synonyms for all categories of ER model elements. This paper will present a complete set of reasoning rules for detection of synonyms, detailed example of an ontology and appropriate CDM diagram with their formalization into Prolog facts, comprehensive description of the reasoning rules application process with examples for each reasoning rules testing, as well as the empirical study that demonstrates applicability of the proposed approach and the developed DMV tool.

The remainder of the paper is organized as follows: the section "Background" describes theoretical concepts and provides a broader context of the research, the section "Related" work presents literature review results, the section "System for detecting synonyms in conceptual data model" presents the proposed approach and the process of synonyms detection with reasoning rules. Section "Example" presents a simple ontology and appropriate CDM, queries and results in extracted synonyms. Section "Empirical Study" presents research methodology and results of testing the proposed approach with students' exams, and section Conclusion gives an overview of the results, application areas, and future research implications.

BACKGROUND

Data quality, as a wide category, includes the quality of data models (Scannapieco, Missier, & Batini, 2005). Entity-relationship (ER) grammar (Chen, 1976) was commonly used in conceptual modeling until the wide application of ontology languages (Bera, Krasnoperova, & Wand, 2010). A general definition of ontology categorizes it as a type of conceptual data model (Spaccapietra, March, & Kambayashi, 2002). Conceptual data models are considered as separate concepts in practice (Weber, 2002), as data models are often task-specific and implementation-oriented, while ontologies are generic and task-independent. (Spyns, Meersman, & Jarrar, 2002).

Ontology could improve the quality of the conceptual model, thereby reducing errors and saving resources in software development (Shanks, Tansley, & Weber, 2003; Keet & Artale, 2008). Ontology can be enriched with synonyms, which broaden the ontology usage possibilities (Kastrati, Shariq Imran, & Yildirim-Yayilgan, 2016). These lexical data are suitable for detecting similarities or synonyms in other types of models that can be mapped with ontology (Kastrati et al., 2016). Such enriched ontology could be used to refine the design of a conceptual model. The inclusion of synonyms in ontology helps to identify the most appropriate label for an entity, attribute, or relationship in CDM, assuming that the terms used in the ontology are the most appropriate (Sugumaran & Storey, 2006).

Synonyms could be defined as a relationship between words usually based on the word sense (not form) and the synonym relations between words are context or domain dependent (Shindo, Fujino, & Nagata, 2010). Synonym-related research is primarily conducted in the area of lexical (textual) analysis in applications such as library management systems (Beal, 2008), semantic web (Mousavi, Gao, & Zaniolo, 2013), and web search engines (Hu, Shao, & Ruan, 2015).

RELATED WORK

Synonym-related research results in the area of data models, databases and data processing are found within broader research related to semantics issues (Naimen, & Ouksel, 1995). Synonym-related issues

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/using-ontology-and-rule-based-reasoning-for-conceptual-data-models-synonyms-detection/230292

Related Content

Rough Sets: A Versatile Theory for Approaches to Uncertainty Management in Databases
Theresa Beaubouefand Frederick E. Petry (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1127-1150).*
www.irma-international.org/chapter/rough-sets-versatile-theory-approaches/7962

Assessment of Students by a Teacher with a Handheld Device and a Networkable Database
C. Paul Newhouse (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1309-1319).*
www.irma-international.org/chapter/assessment-students-teacher-handheld-device/7974

An ER Based Methodology for Modeling User Views and Detecting Derived Relationships
Dinesh Batraand Jeffrey A. Hoffer (1994). *Journal of Database Management (pp. 3-17).*
www.irma-international.org/article/based-methodology-modeling-user-views/51128

Integrating Digital Signatures with Relational Databases: Issues and Organizational Implications
Randal Reidand Gurpreet Dhillon (2003). *Journal of Database Management (pp. 42-51).*
www.irma-international.org/article/integrating-digital-signatures-relational-databases/3294

INDUSTRY AND PRACTICE: Solving the Partitioning Problem in Database Design
Chun Hung Cheng, Chon-Huat Gohand Anita Lee-Post (1999). *Journal of Database Management (pp. 36-38).*
www.irma-international.org/article/industry-practice-solving-partitioning-problem/51211