Ontology Mapping Validation: Dealing with an NP-Complete Problem

Felipe Serpeloni, Unicamp, Brazil
Regina Moraes, Unicamp, Brazil
Rodrigo Bonacin, Center for Information Technology Renato Archer and Public Research Centre Henri Tudor, Brazil

ABSTRACT

The use of ontologies and ontology mappings is increasing in companies. Often the same context is modeled in different ontologies. Mapping is necessary to integrate these ontologies; however, in many cases these mappings are incorrect, i.e., they incorrectly link semantic concepts with different meanings. Tools that validate these mappings are necessary to ensure reliable communication between heterogeneous systems. This validation cannot be done in a completely automatic way, because the mappings are based on human interpretation. This work describes a semi-automatic tool that supports this activity, based on graphs that generate instances validated in a semi-automatic process that aims to ensure mapping robustness. This algorithm deals with an NP-Complete problem in order to generate all the instances. This paper presents a first prototype of the tool and the methodology used to validate the instances automatically generated by the tool.

Keywords: Model-Based Testing, NP-Complete Problem, Ontologies, Ontology Mapping Validation, Robustness Testing

INTRODUCTION

Everyday more and more companies are using ontologies in their information systems, and ontology mappings are very important to make the communication among these systems possible. These mappings always require human intervention, and manual work is still necessary, which can therefore result in errors when defining ontology mappings. In order to guarantee ontology mapping quality, it is mandatory to validate the mappings created.

According to Cardoso’s survey (Cardoso, 2007), people who deal with education and computer software are the most common ontology users. It is important to emphasize that 66% of the people who answered the survey were Academics, and the software system products were the first to use Web Semantic. Ontologies are also used in business services, life sciences, communication, media, and by governments and healthcare providers.

The same survey (Cardoso, 2007) pointed out the main reasons why these organizations had to use ontologies, including: (1) to share common understanding of the information structure among people or software agents, (2)
to enable reuse of domain knowledge and (3) to make domain assumptions clear.

There are several research projects that propose the automatic or semi-automatic generation of ontology mappings. Research in mapping validation is still however, rare, and additional research is necessary.

This paper presents an algorithm and a prototype that implements semi-automatic ontology mapping validation. The objective is to improve the robustness of the ontology mappings. It is proposed to automatically create or select instances for the source ontology through the use of graph algorithms. Many problems involving graphs are NP-Complete, including the problems found in our studies. Therefore, to deal with this complexity, not all the concepts of the ontologies will be validated, and the user will be responsible for choosing which ones are important regarding his/her context. This strategy can result in a feasible processing time, as well as validating the most important concepts according to the user’s criteria.

This paper is organized as follow: The next section presents the definitions of the main concepts used in the paper; the project proposal is then presented in detail; the following section presents the algorithm proposed to walk through the graph and generate the instances; an example that illustrates the tool functionalities is also found; and the final considerations are presented in the last section.

BACKGROUND

Ontology, in Web Semantics, can be understood as an explicit definition of a context, where contextualization is a simplified view of the represented world, a collection of objects of the domain, concepts, and their relations (Jacob, 2003). Ontologies are used in artificial intelligence, information sharing, communication, interoperability, and for the reuse of knowledge domains.

In the context of Web Semantics, ontology is composed of classes, properties, relations, axioms, and instances. Classes represent the concepts of the domain. Properties correspond to the characteristics whose values differentiate instances of the same class. Relations represent the interaction between concepts or classes. Axioms define always true statements in the domain. Instances are the representations of specific elements of the concepts, which is the actual information (Ehrig & Sure, 2004).

In an algebraic definition, an ontology is a pair $O=(S,A)$, where $S$ is the vocabulary (classes, attributes, relations) and $A$ are the axioms (Kalfoglou & Schorlemmer, 2003).

A major problem in the use of ontologies is the variety of existing ontologies. Many of them represent similar domains or intersections between domains, but they are modeled in different ways. This heterogeneity complicates any process that uses more than one ontology, such as software interoperability and semantic search.

Ontology mappings are constructed to deal with this problem. These mappings are rules that associate the concepts of one ontology with another. The mapping process can be defined as: “given two ontologies $A$ and $B$, to map an ontology to another means that for any concept in ontology $A$, try to find a corresponding concept in the ontology $B$, with the same or similar semantic, and vice versa” (Ehrig & Sure, 2004).

In an algebraic notation, the mapping from the ontology $O_1=(S_1,A_1)$ to the ontology $O_2=(S_2,A_2)$ is the transformation $f:S_1 \rightarrow S_2$, where $A_2 \models f(A_1)$. All the interpretations that satisfy the axioms of $O_2$ also satisfy the axioms of the transformed $O_1$ (Kalfoglou & Schorlemmer, 2003).

Several ontology matching systems have been created in previous years. In OAEI-2010 (Euzenat et al., 2010), for example, ASMOV (one of the best results) had 0.61 for precision, 0.65 for recall and 0.63 F-measure. Such measures are still far from the levels required for truly dependable software. Consequently, human intervention and manual work is still necessary. Therefore failures can occur, especially when the size of ontologies increases. For this reason, Verification and Validation (V&V) techniques used in software validation can also be useful for ontology mapping validation.

The Verification and Validation (V&V) processes aim to verify if the system is in agree-
Related Content

An Internet Cost Model, Assignment of Costs Based on Actual Network Use
[www.igi-global.com/article/internet-cost-model-assignment-costs/75529?camid=4v1a](www.igi-global.com/article/internet-cost-model-assignment-costs/75529?camid=4v1a)

Spatio-Temporal Portals for Continuously Changing Network Nodes
[www.igi-global.com/chapter/spatio-temporal-portals-continuously-changing/17991?camid=4v1a](www.igi-global.com/chapter/spatio-temporal-portals-continuously-changing/17991?camid=4v1a)

Social Web for Large-Scale Biosensors
[www.igi-global.com/article/social-web-large-scale-biosensors/75199?camid=4v1a](www.igi-global.com/article/social-web-large-scale-biosensors/75199?camid=4v1a)
Managing Architectural Reconfiguration at Runtime
www.igi-global.com/article/managing-architectural-reconfiguration-runtime/78353?camid=4v1a