M

Mapping Ontologies by Utilising Their Semantic Structure

Yi Zhao
Fernuniversitaet in Hagen, Germany

Wolfgang A. Halang
Fernuniversitaet in Hagen, Germany

INTRODUCTION

As a key factor to enable interoperability in the Semantic Web (Berners-Lee, Hendler & Lassila, 2001), ontologies are developed by different organisations at a large scale, also in overlapping areas. Therefore, ontology mapping has come into forth to achieve knowledge sharing and semantic integration in an environment where knowledge and information are represented by different underlying ontologies.

The ontology mapping problem can be defined as acquiring the relationships that hold between the entities of two ontologies. Mapping results can be used for various purposes such as schema/ontology integration, information retrieval, query mediation, or web service mapping.

In this article, a method to map concepts and properties between ontologies is presented. First, syntactic analysis is applied based on token strings, and then semantic analysis is executed according to WordNet (Fellbaum, 1999) and tree-like graphs representing the structures of ontologies. The experimental results exemplify that our algorithm finds mappings with high precision.

BACKGROUND

Borrowed from philosophy, ontology refers to a systematic account of what can exist or 'be' in the world. In the fields of artificial intelligence and knowledge representation, ontology refers to the construction of knowledge models that specify a set of concepts, their attributes, and the relationships between them. Ontologies are defined as "explicit conceptualisation(s) of a domain" (Gruber, 1993), and are seen as a key to realise the vision of the Semantic Web.

Ontology, as an important technique to represent knowledge and information, allows to incorporate semantics into data to drastically enhance information exchange. The Semantic Web (Berners-Lee, Hendler & Lassila, 2001) is as a universal medium for data, information, and knowledge exchange. It suggests to annotate web resources with machine-processable metadata. With the rapid development of the Semantic Web, it is likely that the number of ontologies used will strongly increase over the next few years. By themselves, however, ontologies do not solve any interoperability problem. Ontology mapping (Ehrig, 2004) is, therefore, a key to exploit semantic interoperability of information and, thus, has been drawing great attention in the research community during recent years. This section introduces the basic concepts of information integration, ontologies, and ontology mapping.

Mismatches between ontologies are mainly caused by independent development of ontologies in different organisations. They become evident when trying to combine ontologies which describe partially overlapping domains. The mismatches between ontologies can broadly be distinguished into syntactic, semantic, and structural heterogeneity. Syntactic heterogeneity denotes differences in the language primitives used to specify ontologies, semantic heterogeneity denotes differences in the way domains are conceptualised and modelled, while structural heterogeneity denotes differences in information structuring.

There have been a number of previous works proposed so far on ontology mapping (Shvaiko, 2005, Noy, 2004, *Sabou, 2006, Su, 2006*). In (Madhavan, 2001), a hybrid similarity mapping algorithm has been introduced. The proposed measure integrates the linguistic and structural schema matching techniques. The matching is based primarily on schema element names, not considering their properties. LOM (Li, 2004) is a semi-automatic lexicon-based ontology-mapping tool that supports a human mapping engineer with a first-cut comparison of ontological terms between the ontologies to be mapped. It decomposes multi-word terms into their word constituents except that it does not perform direct mapping between the words. The procedure associates the WordNet synset index numbers of the constituent words with ontological term. The two terms which have the largest number of common synsets are recorded and presented to the user.

MAIN FOCUS OF THE CHAPTER

Our current work tries to overcome the limitations mentioned above, and to improve precision of ontology mapping. The research goal is to develop a method and to evaluate results of ontology mappings.

In this article, we present a method to map ontologies synthesised of token-based syntactic analysis, and semantic analysis employing the WordNet (Fellbaum, 1999) thesaurus and tree-structured graphs. The algorithm is outlined and expressed in pseudo-code as listed in Figure 1. The promising results obtained from experiments indicate that our algorithm finds mappings with high precision.

Syntax-Level Mapping Based on Tokenisation

Before employing syntactic mapping, a pre-processing is inevitable, which is called tokenisation. Here, ontologies are represented in the language OWL-DL[1]. Therefore, all ontology terms are represented with OWL URI. For example, in ontology "beer", an OWL Class 'Ingredient' is described by

"[OWLClassImpl] http://www.purl.org/net/ontology/beer#Ingredient",

where "*[OWLClassImpl]*" implies OWL class, URL "http://www.purl.org/net/ ontology/beer" addresses the provenance of the ontology, and 'Ingredient' is the class name. Tokenisation should first extract the valid ontology entities from OWL descriptions, which, in this example, is 'Ingredient'.

Moreover, the labels of ontology entities (classes

Figure 1. Pseudo-code of mapping algorithm

```
Input: OWL O1, OWL O2, threshold sigma;
Output: similarity between O1 and O2;
Begin
build tree-structured graphs for O1 and O2, and get their edge sets E1 and E2;
for each child node Ci ∈ E1 do
    for each child node Cj ∈ E2 do
        tokenise Ci and Cj into token sets tci and tcj;
        if (tci unequal to tcj) then
           calculate syntactic level similarity Sim_syn between tci and tcj;
           if (Sim_syn < sigma) then // semantic mapping
              compute semantic-level similarity of tci, tcj based on WordNet;
              if (tci and tcj have no WordNet relationship) then
                 determine similarity Sim_tsg with the specific properties and
                 relationships between their parent/child nodes in ts-graphs
              fi
           fi
        fi
    od
od
end
```

Related Content