Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 504 – 513

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

A method of identifying ontology domain

Dan Wu*, Anne Håkansson

Software and Computer Systems, ICT School, Royal Institute of Technology KTH, Sweden

Abstract

Metadata, such as the domain description and the purpose of an ontology, can be used to describe the context of ontologies for ontology integration. However, these metadata are not always available in ontologies. To solve the problem, a method is developed to automatically discern the domain of an ontology. This method uses a so-called core domain ontology, rules and an ontology reasoner to identify the domain. The core domain ontology is a light weight ontology that consists of the essential concepts of a domain. Rules and the ontology reasoner are used to test if the core domain ontology is consistent with an ontology for which the domain needs to be identified. If the two ontologies are not in violation, then the method confirms the consistency between them, that is, the test ontology shares the same domain as the core domain ontology. If the core domain ontology shows inconsistency with the test ontology, they do not share the same domain and then the ontology can be used to compare with another core domain ontology. Experiments on the core domain ontology for the conference domain show good results. Ten ontologies of mixed domains are compared with the core conference ontology. Eight ontologies' domain are correctly identified, out of which, four ontologies are identified as sharing the same ontology domain.

Peer-review under responsibility of KES International.

Keywords: domain identification of ontology; metadata of ontology and ontology match and integration

1. Introduction

Metadata of ontologies are stored in an ontology repository [1-3] and used to improve the use of ontologies on the semantic web. In our previous research, metadata of ontologies are extracted to support building context rules to provide information for ontology integration [4]. However, many ontologies lack such information [1; 3]. To

* Corresponding author.
E-mail address: dwu@kth.se

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.131

tackle the problem, a method is developed and applied to identify domain information of an ontology. If the domains of ontologies are available, other ontologies that are in the same domain may be used as the contextual information for the semantic ontology integration. Therefore, extra information beyond the integrating ontologies can be used to refine the ontology integration [4].

Context is considered important for building adaptive systems [5;6]. In order to build context and provide contextual information for ontology integration, the metadata of ontologies, such as domain description and purpose, is needed. This paper proposes a method for identifying the domain of an ontology. The following text is organized in six sections: section two describes the related work; section three describes the proposed method, the core domain ontology and the rules used by the method; section four explains the experiment; presents the result and discusses the method and the experiment; and finally, section five concludes the work and points out future work.

2. Related work

To our best knowledge, there is no related work in the area of identifying the domain of an ontology to build context for ontology integration. However, the ontology domain as metadata and algorithms for metadata generating, have been studied in related areas, namely ontology reuse.

Simperl *et al.* [1] investigate how metadata are used for research in ontology engineering, ontology repository and the semantic web tools. The commonly used metadata are the description of ontology domain, availability and licensing conditions, development status, the release date, and also the structure data of ontologies. Two ontology search engines Swoogle[†] and Watson[‡] provide a small amount of ontology metadata, such as URI, ontology format, the size of a file, comments, number of classes, properties and individuals, can be automatically extracted. Beside these metadata, Swoogle extracts data about encoding, ontology ratio, number of statements and information about the ontology discovery process. Watson, on the other hand, shows user reviews, locations and imports. These metadata are quite different from the metadata of ontology domain.

To overcome the difficulty of metadata acquisition, Simpler *et al.*[1] proposed an approach for automatic metadata acquisition, called OMEGA. First, a parser harvests metadata elements from the tags of an ontology. Secondly, several ontology metadata vocabulary (OMV) elements are derived with the help of Google APIs[§], or Wikipedia Categorical Index[**]. Furthermore, an inference engine, such as Pellet[††], can be used to extract metadata automatically. Finally, existing metadata information is reused from online ontology repositories, such as the DAML Library, and search engines Swoogle and Watson.

The OMEGA acquisition tool contains an algorithm that automatically returns *hasDomain*. To discover the domain of an ontology, keywords and key classes are extracted from the ontology, which are compared with DMOZ[‡‡] web category. The weight of a keyword k depends on the highest category level that k appears and it's frequency as the category name. If one keyword appears in sever sub-categories, a maximum likelihood algorithm is applied. For each keyword, the algorithm computes full category paths of the sub-categories that match the keyword, and then it counts the number of distinct top categories to calculate the likelihood of the related domain. The highest likely top DMOZ category based on all the keywords represents the domain. DMOZ is a large and comprehensive human-edited directory of the web, and the domains are represented by words in a hierarchy.

[†] http://swoogle.umbc.edu/
[‡] http://watson.kmi.open.ac.uk/
[§] http://code.google.com
[**] http://en.wikipedia.org/wiki/Portal:Contents/Categorical_index
[††] http://www.mindswap.org/2003/pellet/
[‡‡] http://www.dmoz.org/