Analysis of an Iterative Algorithm for Term-Based Ontology Alignment

Shisanu Tongchim, Canasai Kruengkrai, Virach Sornlertlamvanich, Prapass Srichaivattana, and Hitoshi Isahara

Thai Computational Linguistics Laboratory,
National Institute of Information and Communications Technology,
112,Paholyothin Road, Klong 1, Klong Luang, Pathumthani 12120, Thailand
{shisanu, canasai, virach, prapass}@tcllab.org, isahara@nict.go.jp

Abstract. This paper analyzes the results of automatic concept alignment between two ontologies. We use an iterative algorithm to perform concept alignment. The algorithm uses the similarity of shared terms in order to find the most appropriate target concept for a particular source concept. The results show that the proposed algorithm not only finds the relation between the target concepts and the source concepts, but the algorithm also shows some flaws in the ontologies. These results can be used to improve the correctness of the ontologies.

1 Introduction

To date, several linguistic ontologies in different languages have been developed independently. The integration of these existing ontologies is useful for many applications. Aligning concepts between ontologies is often done by humans, which is an expensive and time-consuming process. This motivates us to find an automatic method to perform such task. However, the hierarchical structures of ontologies are quite different. The structural inconsistency is a common problem [1]. Developing a practical algorithm that is able to deal with this problem is a challenging issue.

The objective of this research is to investigate an automated technique for ontology alignment. The proposed algorithm links concepts between two ontologies, namely the MMT semantic hierarchy and the EDR concept dictionary. The algorithm finds the most appropriate target concept for a given source concept in the top-down manner. The experimental results show that the algorithm can find reasonable concept mapping between these ontologies. Moreover, the results also suggest that this algorithm is able to detect flaws and inconsistency in the ontologies. These results can be used for developing and improving the ontologies by lexicographers.

The rest of this paper is organized as follows: Section 2 discusses related work. Section 3 provides the description of the proposed algorithm. Section 4 presents experimental results and discussion. Finally, Section 5 concludes our work.

R. Dale et al. (Eds.): IJCNLP 2005, LNAI 3651, pp. 346–356, 2005.

2 Related Work

Daudé *et al.* [2] used a relaxation labeling algorithm – a constraint satisfaction algorithm – to map the verbal, adjectival and adverbial parts between two different WordNet versions, namely WordNet 1.5 and WordNet 1.6. The structural constraints are used by the algorithm to adjust the weights for the connections between WN1.5 and WN1.6. Later, some non-structural constraints are included in order to improve the performance [3].

Asanoma [4] presented an alignment technique between the noun part of WordNet and Goi-Taikei 's Ontology. The proposed technique utilizes sets of Japanese and/or English words and semantic classes from dictionaries in an MT system, namely ALT-J/E.

Chen and Fung [5] proposed an automatic technique to associate the English FrameNet lexical entries to the appropriate Chinese word senses. Each FrameNet lexical entry is linked to Chinese word senses of a Chinese ontology database called HowNet. In the beginning, each FrameNet lexical entry is associated with Chinese word senses whose part-of-speech is the same and Chinese word/phrase is one of the translations. In the second stage of the algorithm, some links are pruned out by analyzing contextual lexical entries from the same semantic frame. In the last stage, some pruned links are recovered if their scores are greater than the calculated threshold value.

Ngai *et al.* [6] also conducted some experiments by using HowNet. They presented a method for performing alignment between HowNet and WordNet. They used a word-vector based method which was adopted from techniques used in machine translation and information retrieval. Recently, Yeh *et al.* [7] constructed a bilingual ontology by aligning Chinese words in HowNet with corresponding synsets defined in WordNet. Their alignment approach utilized the co-occurrence of words in a parallel bilingual corpus.

Khan and Hovy [8] presented an algorithm to combine an Arabic-English dictionary with WordNet. Their algorithm also tries to find links from Arabic words to WordNet first. Then, the algorithm prunes out some links by trying to find a generalization concept.

Doan *et al.* [9] proposed a three steps approach for mapping between ontologies on the semantic web. The first step used machine learning techniques to determine the joint distribution of any concept pair. Then, a user-supplied similarity function is used to compute similarity of concept pairs based on the joint distribution from the first step. In the final step, a relaxation labeling algorithm is used to find the mapping configuration based on the similarity from the previous step.

3 Proposed Algorithm

In this section, we describe an approach for ontology alignment based on term distribution. To alleviate the structural computation problem, we assume that the considered ontology structure has only the hierarchical (or taxonomic) relation. One may simply think of this ontology structure as a general tree, where each node of the tree is equivalent to a concept.