SEMCL: A Cross-Language Semantic Model for Knowledge Sharing

Weisen Guo, University of Tokyo, Japan
Steven B. Kraines, University of Tokyo, Japan

ABSTRACT

To promote global knowledge sharing, one should solve the problem that knowledge representation in diverse natural languages restricts knowledge sharing effectively. Traditional knowledge sharing models are based on natural language processing (NLP) technologies. The ambiguity of natural language is a problem for NLP; however, semantic web technologies can circumvent the problem by enabling human authors to specify meaning in a computer-interpretable form. In this paper, the authors propose a cross-language semantic model (SEMCL) for knowledge sharing, which uses semantic web technologies to provide a potential solution to the problem of ambiguity. Also, this model can match knowledge descriptions in diverse languages. First, the methods used to support searches at the semantic predicate level are given, and the authors present a cross-language approach. Finally, an implementation of the model for the general engineering domain is discussed, and a scenario describing how the model implementation handles semantic cross-language knowledge sharing is given.

Keywords: Cross-Language, Knowledge Sharing, Logical Inference, Ontology, Semantic Web, SEMCL

INTRODUCTION

In the age of information explosion, knowledge sharing can significantly increase social capital. The problem of knowledge sharing in the organizations was studied by many researchers of knowledge management field that indicated that the organization culture is more important aspect than technical problems (Widen-Wulff & Suomi, 2003; Widen-Wulff & Ginman, 2004; Tang, 2007; Tang & Zhang, 2008). But the technical problems of knowledge sharing on the Internet are critical issue, especially in the case of much of knowledge is represented in diverse languages that limits our ability to share and search knowledge globally. The traditional approach to share knowledge across diverse languages by manually translating each knowledge resource from the original language to all of the other languages is too slow and costly for tasks such as sharing scientific findings between researchers.

Automated cross-language technologies have been developed that use natural language processing (NLP) technologies to extract keywords for matching knowledge resources...
between different languages. However, NLP-based approaches cannot produce accurate matching results because of the ambiguity of natural language (Hunter & Cohen, 2006). Even thesauri or classification schemata are insufficient (Goldschmidt & Krishnamoorthy, 2008) because they do not support expressions of semantic relationships between keywords or named entities in text. Furthermore, the need to handle multiple languages in cross-language knowledge sharing models exacerbates the problem of natural language ambiguity. Some approaches to decrease the ambiguity have been reported in the literature. For example, Littman et al. (1998) used a latent semantic indexing technique to implement cross-language information retrieval. However, even these sophisticated NLP technologies do not address the fundamental issue of ambiguity in representing knowledge with natural language, an issue that is particularly problematic in a multilingual knowledge sharing situation.

Semantic Web technologies can be used to express knowledge in a computer-interpretable form. They enable matching at a semantic predicate level, e.g., matching of both named entities and predicates stating the semantic relationships between them. Specifically, ontologies constructed in a language such as OWL-DL can represent domain knowledge within a description logic (DL) formalism (OWL, 2004). Then DL-based inference can be used in knowledge search to find more useful matching results (Guo & Kraines, 2008).

We present a cross-language knowledge sharing model in this paper that is based on ontologies and logical inference. Using this model, knowledge providers can publish knowledge resources in their native languages, and knowledge seekers can search for knowledge in different languages, thereby enabling cross-language knowledge sharing. Furthermore, both the descriptors of the knowledge resources and the search queries are represented in a form that can be interpreted semantically by a computer, which enables the computer to infer embedded meaning that is implied but not explicitly expressed. Therefore, the knowledge system implementing this model returns matching results represented in diverse languages that should be more accurate than those of conventional keyword based systems because matching is done at the semantic predicate level.

The rest of the paper is organized as follows. We review the state of the art of knowledge sharing on the Internet. We then present the cross-language knowledge sharing model and describe the cross-language method that we have developed to implement the model. We also discuss the related work.

**KNOWLEDGE SHARING MODELS**

There are many kinds of knowledge in our daily life, and there are many methods for classifying that knowledge. In this paper, we focus on knowledge stored in the Internet, which is explicitly described in formats such as text, pictures, videos, and databases. The Internet environment makes it possible to share knowledge transparently across time and space. However, to make knowledge sharing and searching on the Internet more effective, we need to consider questions such as what knowledge sharing is, what the goal of the knowledge sharing is, and who participates in the knowledge sharing.

Knowledge sharing is an activity through which knowledge is exchanged among people and/or organizations (Lin, 2007). In this paper, we focus on knowledge existing in explicit digital form on the Internet. A knowledge sharing community consists of two main types of the knowledge users: knowledge providers and knowledge seekers. Community members can be both types: each knowledge user may both provide knowledge resources and seek knowledge resources. The goal of a knowledge sharing system is to return the correct knowledge resources to the knowledge seeker. A global-scale knowledge sharing community will invariably include knowledge users from different countries speaking different languages.

We classify the conventional knowledge sharing models into three types: TEM, TEMCL
Knowledge Leakage when SMEs Participate in Supply Chains: What Is It About and How Can It Occur?
www.igi-global.com/article/knowledge-leakage-when-smes-participate-in-supply-chains/160846?camid=4v1a

Identifying Suitable Degradation Parameters for Individual-Based Prognostics
www.igi-global.com/chapter/identifying-suitable-degradation-parameters-individual/69676?camid=4v1a