Chapter 12
KYOTO:
A Wiki for Establishing Semantic Interoperability for Knowledge Sharing Across Languages and Cultures

Piek Vossen
VU University Amsterdam, The Netherlands

Eneko Agirre
EHU, Spain

Francis Bond
Nanyang Technological University, Singapore

Wauter Bosma
VU University Amsterdam, The Netherlands

Axel Herold
BBAW, Germany

Amanda Hicks
BBAW, Germany

Shu-Kai Hsieh
National Taiwan Normal University, Taiwan

Hitoshi Isahara
NICT, Japan

Chu-Ren Huang
Hong Kong University, China

Kyoko Kanzaki
NICT, Japan

Andrea Marchetti
CNR-IIT, Italy

German Rigau
EHU, Spain

Francesco Ronzano
CNR-IIT, Italy

Roxane Segers
VU University Amsterdam, The Netherlands

Maurizio Tesconi
CNR-IIT, Italy

ABSTRACT

KYOTO is an Asian-European project developing a community platform for modeling knowledge and finding facts across languages and cultures. The platform operates as a Wiki system that multilingual and multi-cultural communities can use to agree on the meaning of terms in specific domains. The Wiki is fed with terms that are automatically extracted from documents in different languages. The users can

DOI: 10.4018/978-1-61520-883-8.ch012

modify these terms and relate them across languages. The system generates complex, language-neutral knowledge structures that remain hidden to the user but that can be used to apply open text mining to text collections. The resulting database of facts will be browse-able and searchable. Knowledge is shared across cultures by modeling the knowledge across languages. The system is developed for 7 languages and applied to the domain of the environment, but it can easily be extended to other languages and domains.

INTRODUCTION

This chapter describes the KYOTO system for establishing semantic interoperability for text mining and thus for sharing knowledge across languages and cultures. The system can be used by transnational groups in different languages and cultures with the same domain of interest. KYOTO starts from the assumption that language reflects culture and that the linguistic encoding of knowledge and information is therefore culturally biased. Semantic and cultural interoperability is achieved by defining the words and expressions in each language through a shared ontology. An ontology is a formal, language-independent representation of entities that can be used for inferencing and reasoning.

A Wiki environment will help the users to agree on the meaning of the concepts of interest, to share their knowledge and to relate the terms and expressions in their language to this knowledge. This process is guided by automatic acquisition of terms and meanings from the textual documents provided by the users. The collaborative system will help the users review and edit all acquired information, with a special focus on achieving consensus but also for different views and interpretations across languages and cultures. The users can maintain their knowledge over time and work towards interoperability of terms and language by fine-tuning.

The Wiki environment uses a formal representation for generating knowledge from the conceptual modeling. This representation is language neutral and is not shown to the user directly but can be used by computer software to extract detailed information and facts from a document collection. The extraction process will use the ontological patterns and their relation to the words and expressions in each language so that the information can be interpreted in the same way across these languages and cultures. Likewise, the KYOTO system functions as a cross-lingual and cross-cultural information and knowledge sharing platform.

The system is developed within the KYOTO project (ICT-211423, http://www.kyoto-project.eu/), which is co-funded by the European Union[1] and by (national) funding of Taiwan and Japan. The project started in March 2008 and will end in March 2011. Currently, we completed the specification and design phase and we integrated the first versions of the system components. In the project, we will be working on a restricted set of languages: English, Dutch, Italian, Spanish, Basque, Simplified Mandarin Chinese and Japanese. We will also apply the system to the domain of the environment and specifically to the topic of ecosystem services, a global phenomenon with different linguistic and cultural interpretations. Nevertheless, the system is designed in such a way that it can be used for any language and can be applied to any domain.

The chapter is organized as follows. First, we will describe the situation for the environment domain as a user-case for inter-cultural and cross-lingual information exchange. Next, we will describe the current state-of-the-art in knowledge modeling and information extraction, explaining the short-comings and opportunities. In section 4, we will describe the KYOTO system that we are developing, as a proposal to support the complex