Presentation

A Study on the Best Practice for Constructing a Cross-lingual Ontology

Yi-Yun Cheng
Department of Library and Information Science, National Taiwan University, Taiwan
r02126002@ntu.edu.tw

Hsueh-Hua Chen
Department of Library and Information Science, National Taiwan University, Taiwan
sherry@ntu.edu.tw

Abstract

Ontologies, as the fundamental building blocks for the Semantic Web, are the highest-level classification scheme in the family of Knowledge Organization Systems (KOS). With the emergence of big data, ontologies are one of the keys to unraveling the information explosion problems. Under the big data situation, many language cultures are in a pressing need to construct ontologies. Cross-lingual ontology research has thus become a pivotal concern in this global age. Researchers worldwide try to be interoperable with ontologies written not only in English, but also in other languages. Yet, constructing a cross-lingual ontology can be difficult, and a detailed mapping method is often hard to find. The purpose of this study is to establish a feasible practice on building cross-lingual ontologies. The study will focus on the construction of an English-Chinese ontology from an existing source ontology and a KOS source. This study will also address the synonymy and polysemy problems of the target language (Traditional Chinese).

By adopting a three-phase research design, this study begins with the pretest of our mapping practice on a small ontology—W3C's Semantic Sensor Network Ontology (SSN ontology). This phase is to ensure our SPARQL code to parse all the classes in SSN ontology is feasible. In phase two, we try to map our source ontology—the Semantic Web for Earth and Environmental Terminology (SWEET) ontologies—with the KOS term-lists from National Academy of Educational Research (NAER) in Taiwan. In phase three, we model the mapped English/Chinese ontology in Protégé software to explore the prospect of this method.

The results in phase one shows that our SPARQL code can automatically helped us retrieve all 117 classes in SSN ontology into plain text format in a click, suggesting that our practice is a workable one. In phase two and three, a cross-lingual ontology between English and Traditional Chinese is constructed through the implementation of Protégé. The mapping results between the 3,770 SWEET ontologies classes (in English) and the NAER term-lists (in Traditional Chinese) reveal an accuracy of 80.66% on the exact-match terms, while the Chinese synonyms and related terms expressed by SKOS labels are all proven searchable in our primary evaluation. These promising results demonstrate the feasibility of the practice proposed by this study, and further suggest that such approach is suitable to be adopted by future researchers to model their cross-lingual ontologies.

Keywords: ontologies; cross-lingual ontologies; SWEET ontologies