

Abstract

A variety of controlled vocabularies have been developed for different subject access tools. These controlled vocabularies differ in their subject areas, levels of specificity of concepts, degrees of pre/post-coordination, semantic relationships, and their use of languages. In order to enhance the interoperability between these controlled vocabularies, Zeng and Chan (2003) reviewed a range of methods to establish semantic mapping between different KOS in order to improve their interoperability. However, in the digital world, before establishing semantic mappings between different KOS, it is crucial to make the different KOS into semantic web-enabled formats, and use these formats to represent the KOS and the mappings between them. For this purpose, a number of knowledge representation formats have been developed to encode different types of KOS. These knowledge representation formats vary from traditional library classification representation formats to ontological languages.

With the increasing requirement of establishing semantic mappings between different vocabularies, further development of these encoding formats is becoming more and more important. For this reason, four types of knowledge representation formats were assessed: MARC21 for Classification Data in XML, Zthes XML Schema, XTM (XML Topic Map), and SKOS (Simple Knowledge Organisation System). This paper explores the potential of adapting these representation formats to support different semantic mapping methods, and discusses the implication of extending them to represent more complex KOS.

In this context, several criteria and standards have been identified for this particular study, listed as follows:

1. extensibility: to represent a wide range of KOS;
2. accuracy: to express concept knowledge;
3. capability of converting one format to another;
4. ability to represent mappings between different KOS;
5. supporting tools to query formatted data.

Several issues emerge:

1. XML-based formats are limited and cannot represent some of the more complex thesauri or ontologies and the mappings between them, and therefore RDF-based or XTM-based formats are more appropriate to be extended to encode ontological vocabularies;
2. It is impractical to use only one representation format to encode all the controlled vocabularies, because each has its own structures and syntax. More importantly, different representation formats can be converted into each other depending on the specific requirements.
3. Most of these knowledge representation formats do not provide a specified vocabulary to encode mappings between different KOS. It is therefore necessary to adapt these encoding formats further to develop relevant mapping vocabularies for encoding mapping data, because most of these representation formats have not had relevant schemas to make encoding possible.
4. In the KOS community, there is continuing argument about whether to apply term-based or concept-based representation formats to encode the KOS. Most term-based encoding formats are designated to represent thesauri where the basic description element is based on terms. However, end-users may prefer

to use different KOS as knowledge navigators, which emphasises the need to group relevant terms into a concept and represent a tree of the concepts to the users. Thus, it is important to develop a variety of algorithms and applications to encode KOS in both term-based and concept-based forms. An in-depth usability study on the use of subject access services based on KOS is required.

5. Different representation formats will co-exist for a long time, and there are a number of protocols and applications available to support access to encoded data in different formats. Thus, when developing a terminology mapping service, it is hoped that different formats and protocols can be applied together, and make the use of the most appropriate formats to represent different KOS and the mappings between them.