

Delivering a Terminology Registry

Koraljka Golub
UKOLN - The University of Bath, Bath, BA2 7AY, UK
k.golub@ukoln.ac.uk

Douglas Tudhope
University of Glamorgan, Pontypridd, CF37 1DL, UK
dstudhope@glam.ac.uk

1.0. Introduction

The need for vocabularies such as ontologies, classification schemes and thesauri for information discovery and the Semantic Web has been well recognized and established. Due to a large number of different vocabularies, related services, and a variety of potential applications and contexts, the need has appeared for terminology registries.

A terminology registry lists, describes, identifies and points to sets of vocabularies available for use in information systems and services. It can cover free and publicly available, fee-based and restricted, or organisation-internal vocabularies. The registry allows discovery of suitable schemes for information or, potentially, use, by exposing rich metadata about them for navigation and retrieval. Terminology registries can hold vocabulary level information only, or comprise the member terms, concepts and relationships as well, and also list services based on terminology such as automatic classification, term expansion, disambiguation, translation, semantic reasoning. Registries should, if used as a digital infrastructure service, make their content available for both comfortable human inspection and for machine-to-machine (m2m) access.

There are several registries in its basic form, i.e., only listing vocabularies, such as freely available HILT's (HILT) or commercial Taxonomy Warehouse's (Taxonomy Warehouse). In the United States, National Science Digital Library Metadata Registry (NSDL Registry) started being developed to provide "services to developers and consumers of controlled vocabularies and is one of the first production deployments of the RDF-based Semantic Web Community's Simple Knowledge Organization System (SKOS)".

This study was initiated by UK's Joint Information Systems Committee (JISC) and involves two partners, UKOLN, and University of Glamorgan. Its purpose is to analyse issues related to the potential delivery of a terminology registry as a shared infrastructure service within the UK's further and higher education's information environment. It will consider how a registry might support development of terminology and other services within the context of a services-oriented environment.

2.0. Terminology registry in UK

The overall approach for the study involves identifying relevant information available from prior efforts and project documentation, supplemented by information obtained through consultation with key services, projects and executives across digital library, research and learning domains. The study will describe usage scenarios and use cases, investigate requirements and sustainability, study costs and benefits. The study will also look at organisational questions such as who is to create, maintain and host the content of the registry, in particular as to cooperation with similar registries.

Architectural issues will be explored, in particular the potential for co-ordination of registry efforts within the information environment and across domains. Potential configurations

involve the following: independent registry, coordinated efforts with other UK institutions and registries, such as the metadata (IEMSR) or the services ones (HILT), and coordinated efforts at the international level (e.g., with NSDL Registry).

Registries should make their content available for both machine-to-machine access and comfortable human inspection. Content could include the following layers:

- a) A list of vocabularies described with rich metadata (cf. XMDR (eXtended MetaData Registry (XMDR) Project), NKOS Registry Reference Document for Data Elements (2001)). Different vocabulary types need to be recognized for inclusion, and could range from domain ontologies, core ontologies, thesauri, classification schemes, authority files and synonym rings to lexical databases, encyclopaedias and others.
- b) Member terms, concepts and their relationships.
- c) Services based on the vocabularies. These could include searching, disambiguation, query expansion and reformulation, browsing, automated classification, indexing and social tagging support, mapping between the vocabularies, harvesting, semantic reasoning, text mining, and information extraction.
- d) Services supporting creation and maintenance of vocabularies, including suggestions from text mining and social tags.

A terminology registry could be used for discovering and using relevant terminologies in intelligent retrieval tools, including automated query expansion, mapping and disambiguation tools, both automated and human.

References

1. HILT: High-Level Thesaurus. <http://hilt.cdlr.strath.ac.uk/Sources/vocabulary.html>
2. Taxonomy Warehouse. http://www.taxonomywarehouse.com/cat_include.asp
3. NSDL Registry. <http://metadataregistry.org/>
4. IEMSR: Information Environment Metadata Schema Registry.
<http://iemsr.ukoln.ac.uk/iemsr/>
5. eXtended MetaData Registry (XMDR) Project. <http://www.xmdr.org/>
6. NKOS Registry Reference Document for Data Elements. (2001).
http://staff.oclc.org/~vizine/NKOS/Thesaurus_Registry_version3_rev.htm

Acknowledgement

This work is funded by UK's Joint Information Systems Committee (JISC).