

SKOS for an Integrated Vocabulary Structure

Marcia L. Zeng
Kent State University,
USA
mzeng@kent.edu

Wei Fan
China Academy of Sciences,
China
fanwei@mail.las.ac.cn

Xia Lin
Drexel University,
USA
xlin@drexel.edu

Keywords: SKOS; Chinese Classified Thesaurus; integrated schemes

In order to transfer the *Chinese Classified Thesaurus* (CCT) into a machine-processable format and provide CCT-based Web services, a pilot study has been conducted in which a variety of selected CCT classes and mapped thesaurus entries are encoded with SKOS. OWL and RDFS are also used to encode the same contents for the purposes of feasibility and cost-benefit comparison.

CCT is a collected effort led by the National Library of China. It is an integration of the national standards *Chinese Library Classification* (CLC) 4th edition and *Chinese Thesaurus* (CT). As a manually created mapping product, CCT provides for each of the classes the corresponding thesaurus terms, and vice versa. The coverage of CCT includes four major clusters: philosophy, social sciences and humanities, natural sciences and technologies, and general works. There are 22 main-classes, 52,992 sub-classes and divisions, 110,837 preferred thesaurus terms, 35,690 entry terms (non-preferred terms), and 59,738 pre-coordinated headings (*Chinese Classified Thesaurus*, 2005)

Major challenges of encoding this large vocabulary comes from its integrated structure. CCT is a result of the combination of two structures (illustrated in Figure 1): a thesaurus that uses ISO-2788 standardized structure and a classification scheme that is basically enumerative, but provides some flexibility for several kinds of synthetic mechanisms

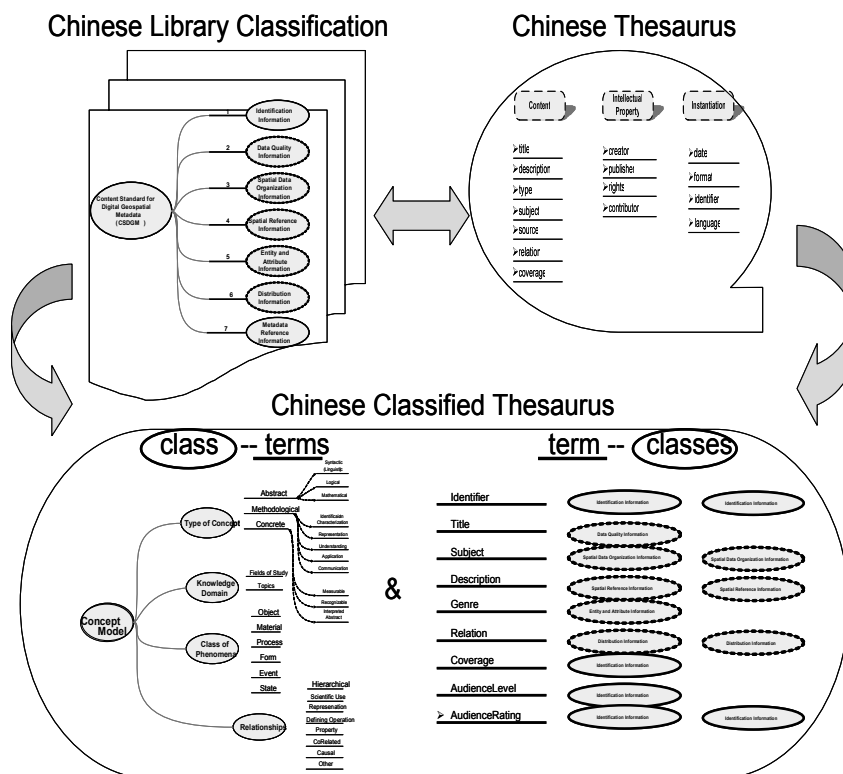


FIG. 1. Illustration of the integrated structure in CCT.

Other challenges include the complex relationships caused by differences of granularities of two original schemes and their presentation with various levels of SKOS elements; as well as the diverse coordination of entries due to the use of auxiliary tables and pre-coordinated headings derived from combining classes, subdivisions, and thesaurus terms, which do not correspond to existing unique identifiers. The poster reports the progress, shares the sample SKOS entries, and summarizes problems identified during the SKOS encoding process. Although OWL Lite and OWL Full provide richer expressiveness, the cost-benefit issues and the final purposes of encoding CCT raise questions of using such approaches.

References:

Chinese Classified Thesaurus. (2005). *National Library of China*. Beijing: Beijing Library Press. Retrieved June 9, 2008, from <http://clc.nlc.gov.cn/ztfzfbgk.jsp>.