Formal Semantic Description Standard Family for Chinese KOS

NKOS Research Office, Shenzhen University, China

(Published on the website of ISO/TC46 Plenary Meeting (Beijing, 2015):
http://tc46.istic.ac.cn/Achievement_ChineseKOS.aspx)

1 Overview

The general objective of the National Social Science Fund project "Research of the Formal Semantic Description Standard Family for Chinese KOS" (Grant No. 12BTQ045) is to propose a formal semantic description standard family for Chinese KOS, which is based on the features and development requirements of domestic KOS, as well as the existing achievements and developing trend of international NKOS (Networked Knowledge Organization System) community. The standard family can provide a series of standard candidates for domestic relevant organizations and individuals to establish various types of NKOS. Furthermore, the supporting systems for the standard family are developed and the technical services are provided, with the aim of promoting the networked development and application of Chinese KOS.

The Formal Semantic Description Standard Family for Chinese KOS consists of three standard members as following:

- Common CNKOS Semantic Description Standard
- OntoThesaurus: a Semantic Description Standard for highly controlled vocabulary
- OWL Semantic Description Standard for Chinese Domain Ontology

Common CNKOS Semantic Description Standard is an extension to SKOS, with a SKOS application specification for Chinese KOS inside. Its predecessor "Semantic Description Standard for Controlled Vocabulary" obtained a grant from the National Library of China, and has been adopted by NLC as a constituent part of its Knowledge Organization Standards. This time, it is improved to Common CNKOS Semantic Description Standard, accompanying with a RDF Schema of CKOS (i.e. the extended vocabulary). This RDF Schema can provide machine-readable definitions of classes and properties in CKOS for applications, in a Semantic Web way.

- SKOS Application Specification for Chinese KOS can transfer Chinese KOS (such as Chinese thesaurus) to a coarse-grained ontology on the Semantic Web with relatively low cost. However, Chinese classification scheme would lose a considerable part of its original semantics if it is modeled only in SKOS. In addition, since SKOS has chosen a relatively loose data model (in order to accommodate low-controlled vocabulary), it doesn't support strict consistency reasoning and evolvement to finer-grained ontology.

- CNKOS Extended Version (SKOS+CKOS) aims to a full description of traditional Chinese controlled vocabulary. Based on the data model of SKOS and

Semantic Web modeling languages RDF/RDFS/OWL, the extensions applicable to classification scheme and/or other types of KOS are put forward. The requirements about checking, reasoning, trimming or expanding are given due consideration. The potential semantics in classification scheme is explicitly and formally represented, so that the automatic classification would be actualized.

OntoThesaurus: a Semantic Description Standard for Highly Controlled Vocabulary is used for highly controlled Chinese KOS, such as Chinese thesaurus, taxonomy and authority, without a loss in their original rigorous semantics. It is based on OWL DL and OWL 2，and can be consistency-checked and reasoned strictly. Dozens of subproperties of broader, narrower or related relationship are proposed, which would lead thesaurus to a finer-grained ontology. Its predecessor Chinese-Thesaurus-Ontology (OntoThesaurus) was supported by another National Social Science Fund project ”Based on ontology and knowledge integration to realize the semantic improvement, sharing and dynamic updating of Chinese thesaurus” (Grant No. 05CTQ001). This time, OntoThesaurus is improved to a description language conforming to the Semantic Web, with an independent namespace and its abbreviation, and strict formal definitions (see RDF Schema of the OntoThesaurus). Meanwhile, a few classes and properties are adjusted according to some new problems founded in the process of CNKOS study.

OWL Semantic Description Standard for Chinese Domain Ontology is a set of reference standards with ease of entry and operation. It is drawn up according to the description requirements involved in the studies of Chinese domain ontology, conducting in advance a comprehensive combing through W3C recommendations of OWL1 and OWL2. It consists of an OWL Guide for Chinese Domain Ontology and the Chinese translations of five W3C recommendations of OWL1 and OWL2 translated by the research team. It is convenient for Chinese domain ontology modelers to reference and use the standard definitions of OWL. These five translations are available on the W3C Website W3C Translations, accessible to those who need to use them around the world.

The KOS represented in SKOS or OWL can be dealt with and supported by existing SKOS or OWL kits in Semantic Web community. The NKOS Research Office of Shenzhen University Library develops two support systems for self-defined CNKOS and OntoThesaurus: the CLSS for classification-like Chinese KOS represented in CNKOS, and the OTCSS for thesaurus-like Chinese KOS represented in OntoThesaurus.

- Support system for CNKOS：Classification Scheme Sharing Service System (CLSS)

(Taking "Chinese Library Classification" Fourth Edition (CLC4) as an example)

CLC4_CLSS Web version: http://www.lib.szu.edu.cn/nkosapp/ProjectForCLC/

CLC4_CLSS Linked Data Service: http://www.lib.szu.edu.cn/nkos/CCT_CLC_V4.0

CLC4_CLSS Web Service

http://www.lib.szu.edu.cn/nkosapp/ProjectForCLC/services/CLCService/getClassEntryType?schemeURI={0}&conceptURI={1}

For implementation technology and invoke methods, please refer to

http://www.lib.szu.edu.cn/nkos/level_2/service/webservice.jsp and

林伟明，曾新红. OntoThesaurus Web Service API 及其应用研究[J]. 图书情报工作. 2010, 54(2): 119-122. (Lin Weiming, Zeng Xinhong. OntoThesaurus Web Service API and Its Application [J]. Library and Information Service, 2010, 54(2): 119-122.)

- Support system for OntoThesaurus：OntoThesaurus Co-construction and Sharing System (OTCSS)

(Taking "Chinese Classified Thesaurus" (CCT) as examples)

CCT1_OTCSS Web version: http://www.lib.szu.edu.cn/nkosapp/ThesaurusProjectForCCTWL/login.jsp

CCT1_OTCSS Linked Data Service: http://www.lib.szu.edu.cn/nkos/CCT_CT_V1.0

CCT2_OTCSS Web version: http://www.lib.szu.edu.cn/nkosapp/ThesaurusProjectForCCT2WL/login.jsp

CCT2_OTCSS Linked Data Service: http://www.lib.szu.edu.cn/nkos/CCT_CT_V2.0

OntoThesaurus-API：Web Service of OTCSS

http://www.lib.szu.edu.cn/nkos/level_2/service/ontothesaurusapi.jsp

2 The application and expansion to ISO25964 and other international standards for resource description

First of all, above research follows existing international standards and corresponding national standards for resource description. However, there are a lot of particularities in Chinese KOS, which need particular specifications or even expansion.

The main standards that this research follows and refers to are as following:

- BSI. ISO 25964-1, Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval [EB/OL]. [2009-12-09]. http://drafts.bsigroup.com/?d=517.
- ISO. ISO/CD 25964-1, Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval[S]. ISO, 2008.
- DCMI. DCMI Metadata Terms [EB/OL]. [2010-03-11].http://purl.org/dc/terms/.
- ANSI/NISO Z39.19-2005, Guidelines for the Construction, Format, and Management of

Monolingual Controlled Vocabularies[EB/OL][S].[2005-12-28]. http://www.niso.org/standards/resources/Z39-19.html
- 中华人民共和国国家标准. GB 13190-91，汉语叙词表编制规则 [S]. 国家技术监督局，1991. (National Standard of the People's Republic of China. GB 13190-91, Guidelines for establishment and development of Chinese thesauri [S]. China State Bureau of Technical Supervision，1991.)
- 中华人民共和国国家标准. GB/T 3860-2008，文献主题标引规则（报批稿）[S]. 国家质量技术监督检验检疫总局. (National Standard of the People's Republic of China. GB/T 3860-2008, Guidelines for determining the subjects and selecting the descriptors (approval draft)[S]. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China.)
- 中华人民共和国国家军用标准. GJB6793- 2009，军用电子分类表编制规则 [S]. 中国人民解放军总装备部. (National Military Standard of the People's Republic of China. GJB6793-2009, Guidelines for establishment of military electronic classification schemes [S]. PLA General Armament Department.)
- W3C. SKOS Simple Knowledge Organization System Reference: W3C Recommendation 18 August 2009 [EB/OL]. [2010-02-23]. http://www.w3.org/TR/2009/REC-skos-reference-20090818/.
- W3C. SKOS Simple Knowledge Organization System Primer: W3C Working Group Note 18 August 2009 [EB/OL]. [2010-02-23]. http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/.
- W3C. OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004[EB/OL]. http://www.w3.org/TR/2004/REC-owl-features-20040210/.
- W3C. OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004[EB/OL]. http://www.w3.org/TR/2004/REC-owl-ref-20040210/.
- W3C. OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004[EB/OL]. http://www.w3.org/TR/2004/REC-owl-guide-20040210/.
- W3C. OWL Web Ontology Language Semantics and Abstract Syntax. W3C Recommendation 10 February 2004[EB/OL]. http://www.w3.org/TR/2004/REC-owl-semantics-20040210/.
- W3C. OWL 2 Web Ontology Language Document Overview: W3C Recommendation 27 October 2009[EB/OL]. [2011-04-13]. http://www.w3.org/TR/2009/REC-owl2-overview-20091027/.
- W3C. OWL 2 Web Ontology Language Quick Reference Guide: W3C Recommendation 27 October 2009[EB/OL]. [2011-04-13]. http://www.w3.org/TR/2009/REC-owl2-quick-reference-20091027/.
- W3C. OWL 2 Web Ontology Language New Features and Rationale: W3C Recommendation 27 October 2009[EB/OL]. [2011-04-13]. http://www.w3.org/TR/2009/REC-owl2-new-features-20091027/.

For example, some content in "11 Facet Analysis" of ISO 25964-1 is applied when the semantic description scheme about "Facet and Node Label" in CNKOS is drawn up, which says:

> It is important to note the pattern of hierarchical relationships around the node labels. Where the label shows the characteristic of division of the superordinate term, all the terms in the

> array following it are true narrower terms of the superordinate term. Thus, in Figure 4, “whole milk”, “buffalo milk” and “sterilized milk” are all narrower terms of “milk”. In contrast, where the label introduces a new facet, the terms that follow are typically not narrower terms of the preceding term: “farm managers” and “cereal products”, etc., are not narrower terms of “agricultural industries”.
>
> Node labels are not thesaurus terms. They are present only for the purposes of systematic display, ……

To correctly model such concept collection structures, SKOS introduces a skos:Collection class. SKOS primer (20090818) suggests two options to represent facet analysis and label node:

1) A set of concepts following a node label can be represented as skos:member of a skos:Collection, the node label is the skos:prefLabel of the skos:Collection, and the collection is defined as a blank node, i.e. no defined URI is allocated. Sometimes it is important to capture the order of concepts in a collection, such as when concepts are listed in alphabetical or chronological order, the skos:memberList and the skos:OrderedCollection can be used in that case. However, according to the SKOS data model, collections are disjoint from concepts. It is therefore impossible to use SKOS semantic relations to have a collection directly fit into a SKOS semantic network. When the concepts in the array following the node label are true narrower concepts of the superordinate concept, the skos:broader or skos:narrower still be used to explicitly assert the hierarchical relationship.
2) One may wonder whether using collections is desirable, as they add complexity to the representations applications have to manipulate. In fact, for some cases, e.g. when KOSs are mainly intended as navigation hierarchies, it seems more intuitive to represent "node labels" or "guide terms" as instances of skos:Concept, and to use normal semantic relationships for linking them to other concepts.

Based on above contents in ISO 25964-1 and SKOS primer, CNKOS suggests: if this kind of concept collection structure appears in Chinese controlled vocabulary, the second way above is preferred, i.e. when the concepts in the array following the node label are true narrower concepts of the superordinate concept, the "node labels" are represent as instances of skos:Concept. Then they can be linked to their superordinate concepts via skos:broader and to their subordinate concepts via skos:narrower. Another case, where the label introduces a new facet, the concepts that follow are typically not narrower concepts of the preceding concept, is not recommended. However, if it exists, the first way above can be adopted.

“OntoThesaurus: a Semantic Description Standard for Highly Controlled Vocabulary” tends to reserve the original hard-earned rigorous semantics of highly controlled Chinese KOS. Strict consistency-checking and reasoning, and evolvement to finer-grained ontology have been got careful consideration.

For example, the hierarchical relationships in the highly controlled vocabulary are transitive, irreflexive and asymmetric. OntoThesaurus explicitly defines the ont:broader and ont:narrower as instances of owl:TransiveProperty, owl: IrreflexiveObjectProperty and owl: AsymmetricObjectProperty. Furthermore, referring to ANSI/NISO Z39.19-2005, three subproperties are defined for ont:broader and ont:narrower respectively: ont:broaderGeneric, ont:broaderInstance, ont:broaderPart, and ont:narrowerGeneric, ont:narrowerInstance, ont:narrowerPart. These extensions also comply with the rules about three types of hierarchical relationship in ISO 25964-1. Corresponding English labels for properties in OntoThesaurus follow the labels in ISO 25964-1, such as UF, BT, BTG, BTI, BTP, NT, NTG, NTI, NTP, TT, RT, SN, et al.

Also referring to ANSI/NISO Z39.19-2005, OntoThesaurus defines 11 pairs of relatively common subproperties for ont:related. Each pair of subproperties is mutually inverse. In order to facilitate human understanding and judgment, these subproperty names keep the name information of two sides of the relationship as much as possible, and the Chinese and English labels for them are also drafted. These subproperties of ont:related can represent those relationships between general “related” and the most specific relationship between two concepts within specific domain. They are relatively common for various fields and can be dealt with in a universal way. If the most specific relationship need to be described, the standard for domain ontology (i.e. the OWL) is recommended.