SKOS for an Integrated Vocabulary Structure

Marcia L Zeng
Kent State University, USA

Wei Fan
Chinese Academy of Sciences, China

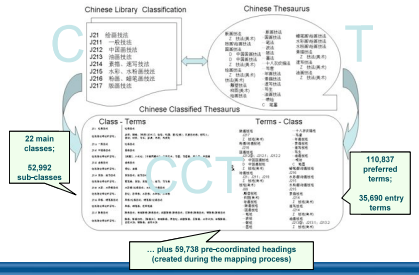
Xia Lin
Drexel University, USA

Introducing CCT

Chinese Classified Thesaurus

1. CCT Vocabulary

- a collected effort led by the National Library of China
- an integration of the national standards *Chinese Library Classification* (CLC) and *Chinese Thesaurus* (CT)
- a manually created mapping product, providing for each of the classes the corresponding thesaurus terms, and vice versa



2. CCT Format

a combination of ...

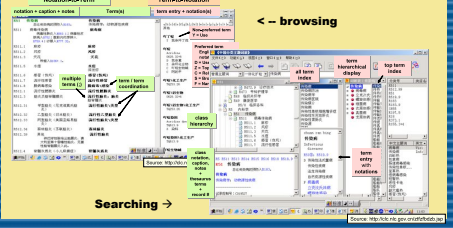
CNMARC Format for Classification Data (a.k.a. CLCMARC) Based on UNIMARC Classification Format & CNMARC for Authorities – Subject Authorities Based on UNIMARC Authorities format

（四）数据区： IV. Data

0– 标识块 Identification Block
1– 编码信息块 Coded Information Block
2– 类目标识块 Heading Block
3– 附注块 Notes block
4– 单纯参照块 See Reference Block
5– 相关参照块 See Also Reference Block
6&7- 类号组配说明块 Number Building Block
700-754 索引款目块 Index Term Block
8– 信息来源块 Source Information Block
9– 国内使用块 National Use Block

0– 标识块 Identification Block
1– 编码信息块 Coded Information Block
2– 标目块 Heading Block
3– 标目附注块 Information Note Block
4– 单纯参照框查块 See Reference Tracing Block
5– 相关参照框查块 See Also Reference Tracing Block
6– 分类号块 Classification Number Block
7– 生技标目块 Linking Heading Block
8– 来源信息块 Source Information Block
9– 国内使用块

3. Current CCT Online Interface



SKOS for [enumerative] Classification Systems

Beyond thesauri, there are issues:

1. The order of main classes/ schedules

- Use skos:OrderedCollection to include main classes and used skos:memberList to show the member in an order.

2. Alternative classification notations

e.g., [Q89] environmental biology
Preferred class: X17

Discussion of options:

? skos:notation
Yes – if SKOS allows for more than one notation for one concept.
But how can you show which one is preferred and which one is not?

? skos:altLabel
No. This is not an alternative label. It is a concept, with its own semantic relations.

3. Top Concept types...for auxiliary tables, etc.

- Add local attributes?
e.g., "hasTopTableConcept"

my:hasTopFamilyConcept
my:hasTopClassConcept
my:hasTopTableConcept
… …

4. Notations are constructed in various ways...

From CCT Format -- field 260:

260 分类号（必备，不可重复）
指示符1：类号的编制制度
0 层累制
1 八分法
2 双位制
3 借上级类号
4 借下级类号
5 借同级类号
6 借0编号
7 使用"-"列举编号
8 顺序制编号(例如:A/B/C/D…)
n 以上都不适用的

260 Notation (Mandatory, non-repeatable)
Indicator 1: Notation forming rule
0 hierarchical notation
1 expanding at number 9
2 double-digits
3 borrowing super-ordinate class's notation
4 borrowing sub-ordinate class' notation
5 borrowing co-ordinate class' notation
6 borrowing "0" in notation
7 using "-" for summary number span
8 sequential notations
n other rules

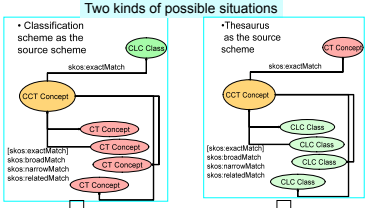
skos:notation doesn't record how a classification notation is built.

SKOS for Mapped Vocabularies

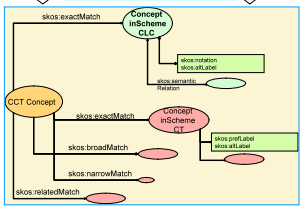
Option 1. CCT as a mapping result

Advantages:
Semantic relationships are clear
Avoids semantic conflicts in applications
Disadvantages:
Complicated
Time-consuming

Two kinds of possible situations



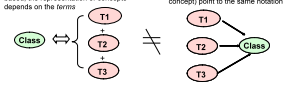
One presentation for both



Sub-issue 1. Many-to-one mapping

multiple terms for one class

• Although the mapping is concept-based, the representation of concepts depends on the *terms*



The result could be: more than one term (each represents a different concept) point to the same notation

Option discussed: deal with a combined term as a **string** only, each representing a concept that did not exist in the original source schemes.

There might be problems for these new strings to be semantically linked with the concepts that each unit of the string originally represents.

Sub-issue 2: One–to -many mapping

CCT specifies degrees of mapping:
- Major mapping
- Minor mapping
- Alternative mapping
- Overlapping

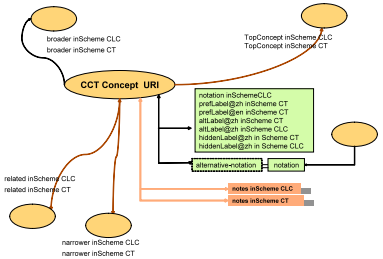
How to indicate the degrees of mapping?
SKOS' 'mapping relation' properties :
- broadMatch
- narrowMatch
- relatedMatch

These are not the same as CCT's mapping relations.

Sub-issue 3: Degree of mapping

Option 2. CCT as a new vocabulary

Treat newly coordinated terms and notations as labels of concepts. (They did not exist in the original source vocabulary.)



Other issues:

- **synthesized numbers (and terms)**

- **the *add* or *divide like* numbers**
  - **i.e., classification number constructed by adding numbers from other parts of the schedule, from a table, or by basing it on a pattern defined in another part of the schedule.**
- **parallel schedules**
  - **e.g., the 'law' class has two parallel schedules.**
- **faceted scheme features**
  - **e.g., the 'bridges' class has three facets; subgroups are listed under node labels.**
- **full, abridged, and extended (+) numbers**

吖啶
Acridine
0625, 32+6

Acknowledgements

This research is the result of collaboration with the Editorial Office of Chinese Library Classification at the National Library of China, especially Dongbo Wang and Shuqing Bu.