ISO 25964: Thesauri and Interoperability with Other Vocabularies

+ IFLA Guidelines for Multilingual Thesauri
+ A KOS Resource Application Profile

Marcia Lei Zeng
Kent State University
Some slides are prepared by Stella Dextre Clarke, leader of the ISO 25964 Working Group

CENDI/NKOS Workshop, National Agricultural Library, MD. October 22, 2009

Current status of ISO 25964

ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
  - Draft finished, to be on ballot soon.
- Part 2: Interoperability with other vocabularies
  - Started after June 2009 meeting in London

- 15 countries participate: Bulgaria, Canada*, China, Denmark*, France*, Germany*, Finland, Korea, New Zealand, South Africa, Spain, Sweden, UK*, Ukraine, USA*
- * active Working Group members 2008-2009

Relationship with older ISO standards

- Will cover and replace two:
  - ISO 2788-1986 *Guidelines for the establishment and development of monolingual thesauri*
  - ISO 5964-1985 *Guidelines for the establishment and development of multilingual thesauri*

What distinguishes ISO 25964-1 from ISO 2788/5964?

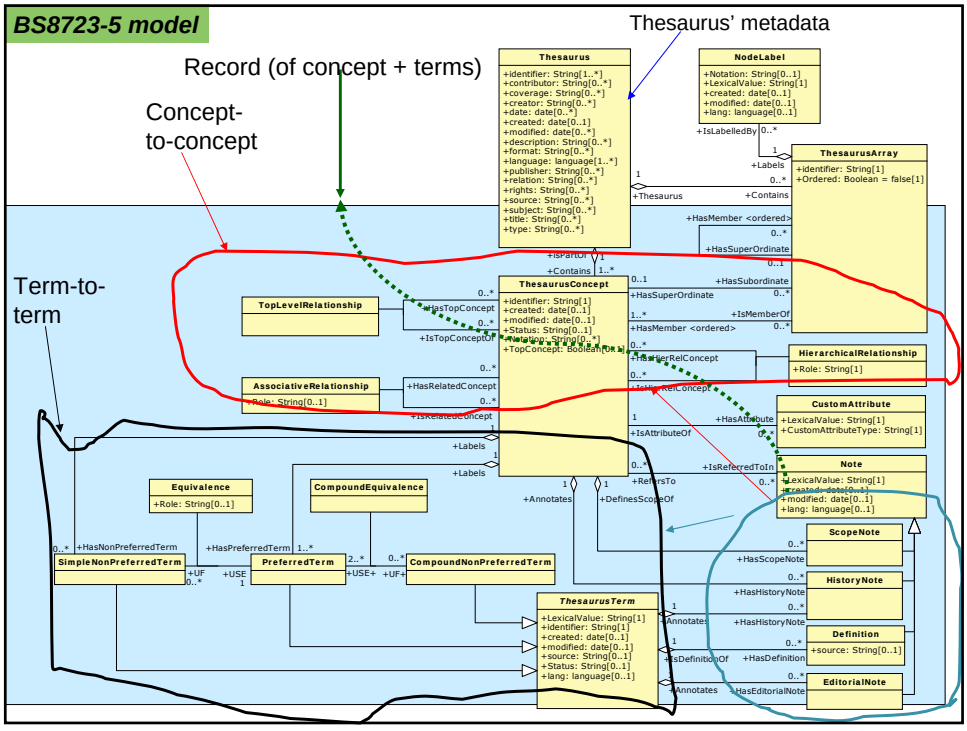
- Clearer differentiation between terms and concepts
- Clearer guidance on applying facet analysis to thesauri
- Some changes to the 'rules' for compound terms
- More guidance on managing thesaurus development and maintenance
- Functional specification for software to manage thesauri
- Data model and XML schema for data exchange
- General overhaul in all areas, e.g. sweeping update of multilingual examples

Relationship with current BS standard

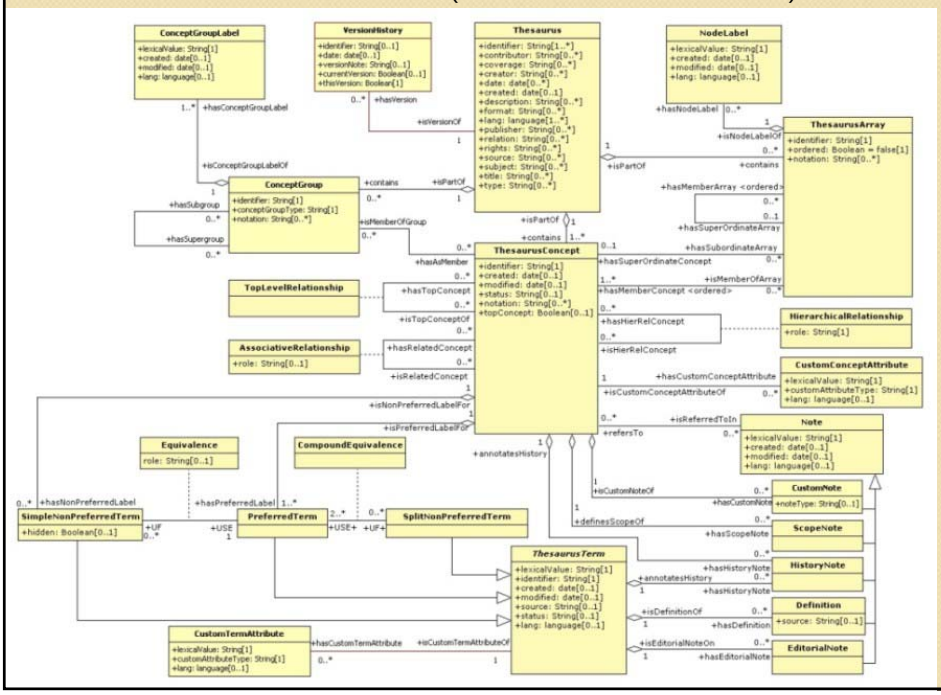
BS 8723: *Structured vocabularies for information retrieval – Guide*

Published 2005-2008

- Part 1: Definitions, symbols and abbreviations
- Part 2: Thesauri
- Part 3: Vocabularies other than thesauri
- Part 4: Interoperability between vocabularies
- Part 5: Exchange formats and protocols for interoperability



ISO 25964 Data model (based on the BS 8723-5 model)



What distinguishes ISO 25964-1 from SKOS?

- ISO 25964-1 emphasize on how to build and manage a vocabulary
- SKOS emphasize on how to publish a vocabulary in a machine-processable format.
- The data models for both standards are concept-based. At a simple level it is easy to convert between them.
- ISO 25964 model provides for all features of a thesaurus
- SKOS model aims to serve several different vocabulary types. It does not provide for some thesaurus features.

Issues for Part 2 (ISO 25964-2)

1. How much description of vocabularies other than thesauri?
2. Whether and how to include "non-symmetrical multilingual thesauri"
3. To provide for interoperability, what do we need in the way of data modeling, exchange formats and protocols?

Issue 1: Coverage (as planned)

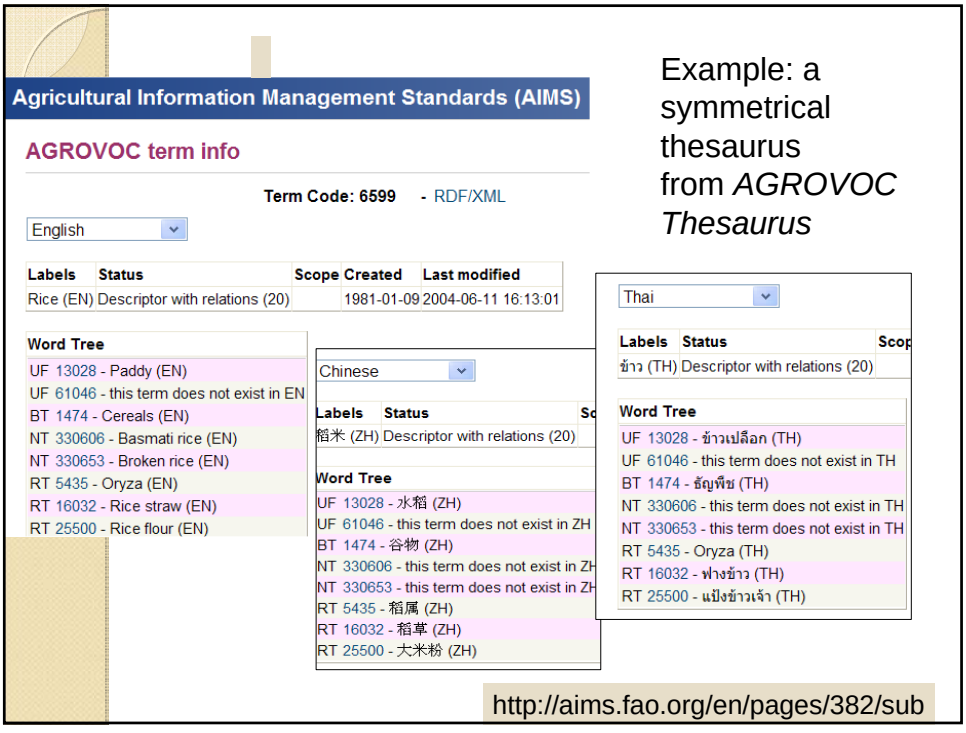
Instead of defining each type of vocabulary and provide guidance of its construction, Part 2 focuses on the interoperability between thesauri and other types of vocabularies.

Vocabulary types include those covered by BS 8723 (2005-2008) *Structured vocabularies for information retrieval – Guide* and NISO Z29.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*

- Classification schemes
- Business classification schemes for records management
- Taxonomies
- Subject heading schemes
- Ontologies/Topic maps/Semantic networks
- Terminologies/Term banks
- Name authority lists
- Synonym rings

Issue 2. non-symmetrical multilingual thesaurus

- A multilingual thesaurus in which the languages do not always share the same structure of hierarchical and associative relationships
- And what is the difference between a non-symmetrical thesaurus and two or more monolingual thesauri with mappings between them?

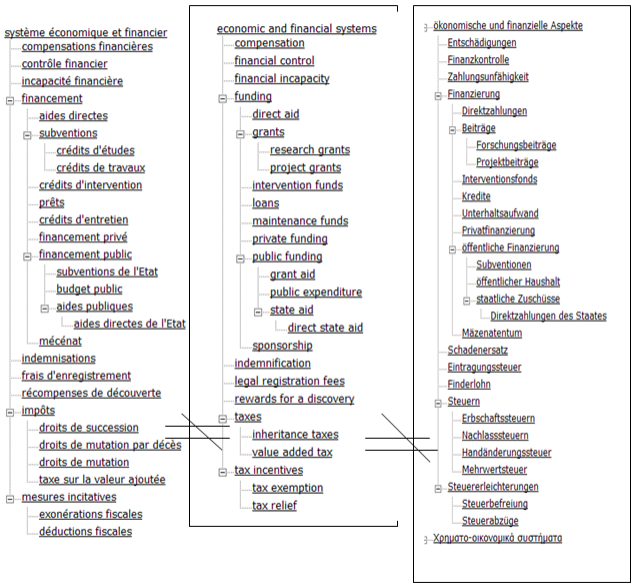


Example: a symmetrical thesaurus from *AGROVOC Thesaurus*

http://aims.fao.org/en/pages/382/sub

French English German

Extracted from: http://www.european-heritage.net/sdx/herein/thesaurus/introduction.xsp

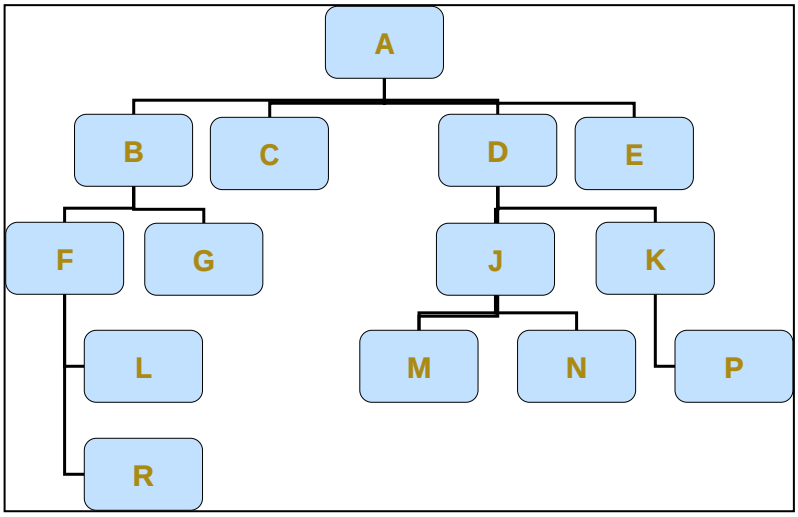


Example: a non-symmetrical situation

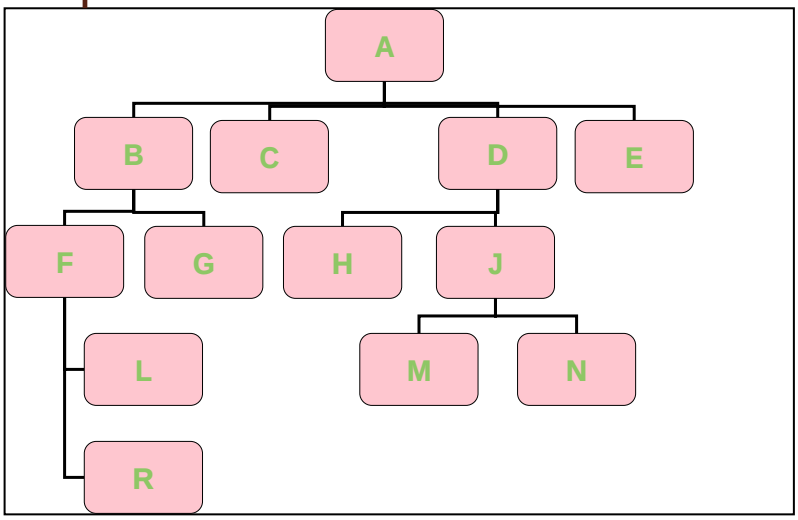
Example: a non-symmetrical situation
Using economics concepts as an example

English	German	French
• taxes ◦ inheritance taxes ◦ value added tax	• Steuern ◦ Erbschaftssteuern ◦ Nachlasssteuern ◦ Handänderungssteuer ◦ Mehrwertsteuer	• impôts • droits de succession • droits de mutation par décès • droits de mutation • taxe sur la valeur ajoutée

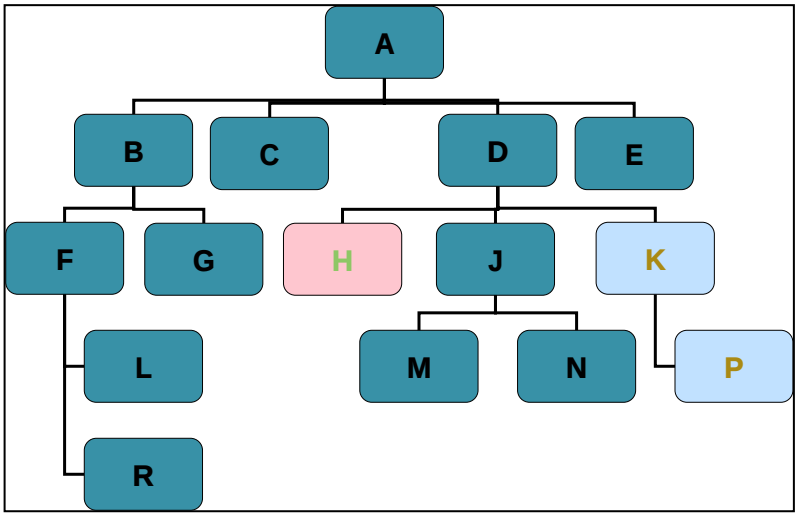
non-symmetrical thesaurus example – Greek in blue



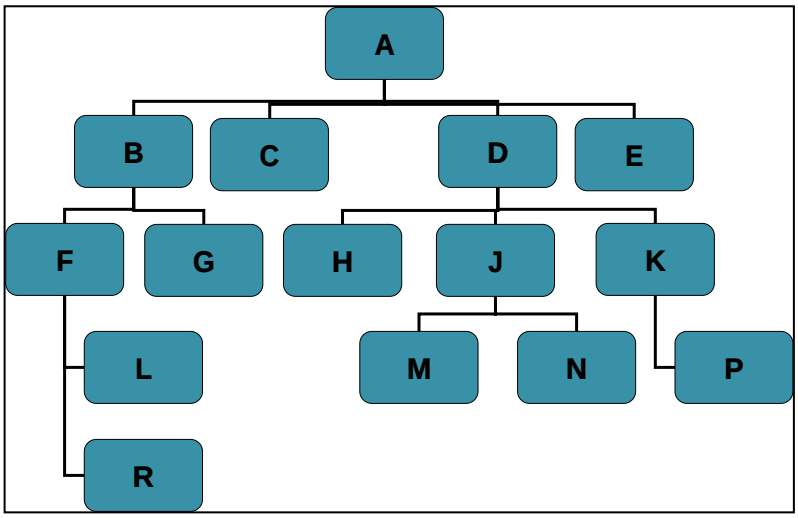
non-symmetrical thesaurus example – Spanish in red



non-symmetrical thesaurus example – teal where structure coincides



Compel both languages to accept loan terms, and you make it symmetrical!



Pros and cons of a non-symmetrical thesaurus

- Each language version retains its own individuality
- Less need to accept artificial terms
- Limitations on interoperability
- Does not conform to the same data model
- Hard to find software to manage all the language versions in concert

Issue 3. What needs for data models, formats and protocols?

- A unified model to cover all vocabulary types?
- A model for each type of vocabulary?
- A syntax for specifying mappings between pairs of vocabularies?
- What are the use cases?

Current plan for ISO 25946 –part 2

- Structural models for interoperability across vocabularies
- Mappings across vocabularies
- Establishing equivalence in practice
- Influence of the application for which mapping is intended
- Managing mappings data
- Display of mapped vocabularies
- Mapping system functionality
- Exchange formats for mappings
- (see next slide) . . .

For each type of vocabulary, at least cover:

- Introduction
- General description
- Scope and role in information retrieval
- Historical note about origin
- Vocabulary control
- Types of xxx
- Semantic components and relationships
- Choice of concepts and terms
- Relationships
- Presentation (when appropriate include info about navigation and searching options)
- Management aspects

International Federation of Library Associations and Institutions

IFLA Guidelines for Multilingual Thesauri

-- Published in 2009 after a world-wide review and revision

Three approaches in the development of multilingual thesauri:

1. building a new thesaurus from the bottom up
   - starting with one language and adding another language or languages
   - starting with more than one language simultaneously
2. combining existing thesauri
   - merging two or more existing thesauri into one new (multilingual) information retrieval language to be used in indexing and retrieval
   - linking existing thesauri and subject heading languages to each other; using the existing thesauri and/or subject heading languages both in indexing and retrieval
3. translating a thesaurus into one or more other languages

Contents covered by the guidelines

- Building multilingual thesauri starting from scratch
  - Structure
  - Morphology and Semantics
- Starting from existing thesauri
  - Merging
  - Linking
- Glossary
- Appendix:
  - *An example of a non-symmetrical thesaurus*

Metadata for KOS Resources

-- Developing a DCMI/NKOS Application profile

A new proposal by NKOS

A review of metadata for KOS resources

(Last year NKOS Workshop at the World Bank)

Cases

- SchemaWeb; Swoogle; BioPortal;11179 Data Element Registries

Standards

- *ISO 11179-2 Information Technology – Metadata registries (MDR)*- Part 2 Classification

Rationale

Basically, metadata for KOS resources will …

- describe specific characteristics of KOS resources
- assist in the discovery of KOS resources
- facilitate the evaluation of the KOS resources for a particular application or use
- facilitate sharing, reusing, and collaboration

Metadata for KOS resources are important to:

- Terminology registries, Service registries, Vocabulary users

Currently there is no standardized metadata element set

NKOS Group's Efforts: 1. KOS Attributes

NKOS Registry - Draft Set of Thesaurus Attributes

(based on Controlled Vocabulary Registry developed by Linda L. Hill and Interconnect Technologies in 1996, with modification led by Gail Hodge)

http://nkos.slis.kent.edu/Thesaurus_Registry.html

Terminology Registry Scoping Study (TRSS), 2008

(PIs: Kora Golub, Doug Tudhope, Trss Final Report to JISC, UK.)

http://www.ukoln.ac.uk/projects/trss/

NKOS Registry, Version 3 with Reference Document for Data Elements

For use with Dublin Core

- core elements only
- consistent with Dublin Core elements and attributes for each element

(Draft developed by Diane Vizine-Goetz, Last updated: 2008

http://nkos.slis.kent.edu/registry3.htm

Version 3

for facilitating the discovery of KOS resources (DC-based) :

- KOS Title (R)
- Alternative Title (O)
- Creator (O)
- KOS Subject (R)
- Description (O)
- Publisher (O)
- Date (R)
- **KOS Type** (R)
- Format (R)
- Identifier (O)
- Language (R)
- KOS Relation (R)
- Rights (O)

for recording specific characteristics, to facilitate the evaluation of the resource for a particular application or use:

- Entity Type (R)
- Entity Value (O)
- Relationships (R)
- Information Given (O)
- Arrangement (R)
- Application (O)

- Minor Subject (O) [Should this be a qualifier of KOS Subject?]

http://nkos.slis.kent.edu/registry3.htm

NKOS Group's Efforts: II. KOS Typology

Taxonomy of Knowledge Organization Sources / Systems
(Gail Hodge et al. 2000--)
http://nkos.slis.kent.edu/KOS_taxonomy.htm

A tentative typology of KOS
(Doug Tudhope, 2006 NKOS Workshop)

http://nkos.slis.kent.edu
(NKOS Website, announcements, listserv, workshop materials since 2000, work-in-progress, etc.)