Knowledge Representation Resources

Sue Ellen Wright
Kent State University, Kent, Ohio
NKOS-CENDI September 2008
Prescribed Parameters

- The range of “KOS” types
- Current standards and what is going on in some of these areas
- Efforts at harmonization & coordination
  - Issues in leveraging across system boundaries and between communities of practice
  - Examples: ISO TC 37 DCR, Terminology, SKOS, OWL-DL
Steps towards a Typology

- **Hodge, Gayle. 2000; 2004.**
  - Taxonomy of Knowledge Organization systems.

- **Soergel, Dagobert. 2001**
  - Evaluation of Knowledge Organization Systems (KOS): Characteristics for describing and evaluating KOS.

- **Wright, Sue Ellen. 2006.**
  - A Typology for Knowledge Representation Resources, ISKO, Vienna

- **Tudhope, Doug. 2006.**
  - A Tentative Typology of KOS: Toward a KOS of KOS?

- **Hlava, Marjorie M.K. 2007.**
  - Insuring Compatibility and Crosswalks.
Steps towards a Typology

**Hodge, Gayle. 2000; 2004.**
- Taxonomy of Knowledge Organization systems. [http://nkos.slis.kent.edu/KOS_taxonomy.htm](http://nkos.slis.kent.edu/KOS_taxonomy.htm)

**Soergel, Dagobert. 2001**
Steps towards a Typology

- Wright, Sue Ellen. 2006.
  - A Typology for Knowledge Representation Resources, ISKO, Vienna

- Tudhope, Doug. 2006.
  - A Tentative Typology of KOS: Toward a KOS of KOS? www.ukoln.ac.uk/nkos/nkos2006/presentations/tudhope.ppt

  - Insuring Compatibility and Crosswalks.
    www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2007/presentations/NKOS%202007-HLava.ppt
TC 37 concept system:
- set of concepts (3.2.1) structured according to the relations among them

Knowledge Organization Schemes
- “a set of concepts, optionally including statements about semantic relationships between those concepts”
- KOS under this definition potentially includes both systematic and non-systematic resources.

Resources with latent concept systems
Knowledge Representation Resources (KRRs)

- Resources that are in themselves well “organized” (systematic)
- Resources that contain latent organizational elements (non-systematic)
- Resource that contains knowledge we can:
  - Manipulate
  - Mine or use to enrich other resources
  - Analyze and reuse (leverage)
  - Use to interact with various tools, either based on common environment planning (or not)
Principle of Systematicity

- Systematic resources include explicitation of relationships (parent-child, meronymy & metonymy, sequentiality, defined edges in logical triads (RDF), etc.)
- Non-systematic resources are not ordered or are conventionally ordered (alphabetical dictionaries, non-mnemonic numerical sequences, etc.)
- Some resources (terminologies, metadata registries [MDRs], etc.) are manifested in both variations
- Degrees of systematicity
Colors

- **Blue**: systematic, represents shallow to deep semantic structures
- **Red**: non-systematic, primarily lists with random or conventional (e.g., alphabetical) ordering principles
- **Green**: hybrid superordinate nodes with both systematic and non-systematic children; texts of various kinds
- **Purple**: WordNet: internally hybrid system; shallow systematics, lexicographical approach
Communities of Practice

- Different communities of practice (CoPs) using different terms for the same concepts and the same terms for different concepts
- KOS defined differently by different CoPs
- Critical concepts subject to indeterminacy
  - *e.g.*, *term*, *terminology*, *metadata*, *ontology*
- Result: indeterminacy in the form of hidden polysemy and synonymy
Language-Purposed Vocabulary: terminology

Terminology *(TC 37 for Terminology)*

- Definition: Set of designations belonging to a special language (reflecting the concepts used in that special language)
- Terminologies (and glossaries) document the use of these designations in discourse, i.e., *uncontrolled vocabulary* used for communication rather than their use for documentation and retrieval as part of a *controlled vocabulary*.
- They can contain concept systems.
- They can also be used to enrich other resources.
thesaurus-related *term*:

- One or more words designating a concept
- A descriptor in a controlled vocabulary used for information management & retrieval
- Thesaurus developers tend to use *terminology* as a synonym for *thesaurus*, whereas terminologists consider terminologies and even terminological concept systems to be different from thesauri.
- This distinction led to a hypothesis: need to identify crosswalk nodes between the systems in order to achieve interoperability.
Terminologies

- **SKOS**: Subject-purposed vocabularies = Subject Language Terminologies
  - Svenonius: SLTs
  - E.g., thesauri: controlled vocabularies
  - Information storage & retrieval

- **TBX**: Language-purposed vocabularies
  - Tudhope
  - Terminological Databases (TDBs)
  - Uncontrolled vocabularies; discourse oriented

- Issues involving mapping data elements between the two CoPs
Hypothesis: 
TDB Term = SKOS Term (Label)?

<table>
<thead>
<tr>
<th>SKOS classes and properties</th>
<th>12620 data category → SKOS</th>
<th>12620 data category ≠ Not in SKOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labels and Terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label</td>
<td>N/A</td>
<td>term (A.1)</td>
</tr>
</tbody>
</table>

Svenonius:

- SLTs designed for the special purpose of retrieving information; extension = all documents about a subject (e.g., all documents about butterflies)
- TDBs designed to document terms used in discourse; extension = the class consisting of all objects covered by the concept (e.g., all butterflies)
Hypothesis: Definition = Definition?

<table>
<thead>
<tr>
<th>Definitions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>definition</td>
<td>definition (A.5.1)</td>
<td>(See discussion)</td>
</tr>
<tr>
<td>scopeNote</td>
<td>explanation (A.5.2)</td>
<td>(See discussion)</td>
</tr>
</tbody>
</table>

- SKOS definition: A *statement or formal explanation of the meaning of a concept*.
- SKOS scopeNote: A *note that helps to clarify the meaning of a concept*.
- These declarations coincide ostensibly with TBX /definition/ and /explanation/, respectively.
Problems with Proposed Solution

- Skos examples:

  `<skos:Concept rdf:about="http://my.example.org/GCL/702#scopeNote">`  
  `<skos:prefLabel xml:lang="en">Competitiveness</skos:prefLabel>`  
  `<skos:scopeNote xml:lang="en">The ability of businesses to compete in local, national or international markets.</skos:scopeNote>`  
  `<;/skos:Concept>`

  `<skos:Concept rdf:about="http://www.example.com/concepts#banana">`  
  `<skos:prefLabel xml:lang="en">banana</skos:prefLabel>`  
  `<skos:definition xml:lang="en">A long curved fruit with a yellow skin and soft, sweet white flesh inside.</skos:definition>`  
  `<;/skos:Concept>`

- Problem: Both of these items are formal definitions in the sense of TBX (ISO 704)
<skos:definition>A feature type category for places such as the Erie Canal</skos:definition>
<skos:scopeNote>Manmade waterway used by watercraft or for drainage, irrigation, mining, or water power</skos:scopeNote> (Mikhalenko 2005)

- This approach, if followed, would suggest flipping the crosswalk assignment, with TBX /definition/ mapping to SKOS scopeNote, and TBX /explanation/ mapping to SKOS definition!
Standards for Language & Knowledge Organization

- Wide range of subject areas
- Focus on the so-called “language industry,” much of which is potentially interesting for information science perspectives
- Wide range of Communities of Practice
- Wide range of standardizing & authoritative bodies
- Overlapping interests & projects
Language and Knowledge Standards
Standards Bodies

- ISO family of standards bodies
- National bodies (NISO, DIN, BSI, etc.)
- Web-oriented standards bodies (W3C, IETF, Unicode, etc.)
- Industry standards (OMG, OASIS, LISA, etc.)
- Professional organizations (ATA, FIT, etc.)
- Research groups and grant teams
- Others?
Thesaurus & Indexing Standards

ISO
- ISO 2788: Monolingual Thesauri
- ISO 5964: Multilingual Thesauri
- ISO 5963: Documents, Subjects & Indexing
- ISO 999: Content, Organization & Presentation of Indexes

ANSI-NISO
- NISO Z39.19 Monolingual Thesauri
- NISO Z39.19-200x: Monolingual Controlled Vocabularies
- NISO TR02-1997: Indexes and Information Retrieval Devices

BSI Structured Vocabularies
- BS 8723-1: Guide. Definitions, etc.
- BS 8723-2: Guide. Thesauri
- BS 8723-3: Guide: Vocabularies other than Thesauri
- BS 8723-4: Interoperability between vocabularies
Current Projects

- Completion of the TC 37 Data Category Registry
- Provision for flexible RDF output from the DCR to support interchange and interoperability
- Experimentation with RDF-based views of critical standards, such as TBX
Data Category Specification

Description Section

Language Section

Name Section

Definition

Example

Data Element Name

Explanation

© Sue Ellen Wright

30 of 34
Data Category Specification
Linguistic Section

- Complex Data Category
- Simple Data Category
- Closed Data Category
- Open Data Category

- Constrained Data Category
- Schema Specific Domain
- Open Conceptual Domain
- Profile Value Domain

- Conceptual Domain
- Value Domain
- Closed Linguistic Section
- Constrained Linguistic Section

Example
Explanation

Example:

Explanation:

Example:

Explanation:

Example:

Explanation:

Example:

Explanation:
ISO TC 37 Terminology and Other Language and Content Resources

Data Category Registry
defining widely accepted linguistic concepts

ISO 12620 provides a framework for defining data categories compliant with the ISO/IEC 11179 family of standards. According to this model, each data category is assigned a unique administrative identifier, together with information on the status or decision-making process associated with the data category. In addition, data category specifications in the DCR contain linguistic descriptions, such as data category definitions, statements of associated value domains, and examples. Data category specifications can be associated with a variety of data element names and with language-specific versions of definitions, names, value domains and other attributes.

This new DCR implementation is currently under development. However, you’re very welcome to give it a try. Notice, that the ISOcat web interface uses a state-of-the-art web interface toolkit which needs a modern browser. For day to day DCR usage the Syntax web interface should still be used.

We welcome any feedback (bug reports, feature requests, etc.). You can use the following methods to contact us:
- file a bug report in Bugzilla (coming soon)
- send an email to the ISOcat system administrator

- Alpha/beta test due Oct. 2008
- Planned as open source software
- Special features:
  - ISO balloting procedures
  - Multilingual solutions
  - Metadata Registry
In the spirit of ISO 11179
OWL-DL Representation of TBX
Navigating Relation Registries

Relation registries

Data category registries

Linguistic resources

MPI

DCR

ISO

DCR

TDS

RR

RR

MPI archive

resource

TDS database
For More Information

- Sue Ellen Wright
  Institute for Applied Linguistics
  Kent State University
  109 Satterfield Hall
  Kent, Ohio 44242, USA

sellenwright@gmail.com