Types of mapping recommended in ISO 25964, and the question of reciprocity

Stella G Dextre Clarke
Project Leader, ISO NP 25964
Overview

- What is ISO 25964? (briefly)
- Recommended types of mapping
- Brief comparison with SKOS
- Reciprocity: when/how do mappings work in both directions?
What is ISO 25964?

ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
- Part 2: Interoperability with other vocabularies

- It updates ISO 2788 and ISO 5964
- Part 1, to be published early 2011, covers monolingual and multilingual thesauri
- Part 2, to be published in 2012, covers mapping between thesauri and other types of vocabulary
- Information retrieval seen as main application; mapping applies to index terms or to search terms
ISO 25964-2 mapping types

- Basic mapping types:
  - Equivalence
  - Hierarchical
  - Associative

- Equivalence mappings can also be marked as “Exact” or “Inexact”
ISO 25964-2 mapping types with examples

- Basic mapping types:
  - Equivalence: Laptop computers EQ Notebook computers
  - Hierarchical: Roads NM Streets; Streets BM Roads
  - Associative: e-Learning RM Distance education

- “Exact” or “Inexact” equivalence
  - Aubergines =EQ Egg-plants
  - Horticulture ~EQ Gardening
Subdivisions of ISO 25964-2 mapping types

- Basic mapping types:
  - Equivalence
    - Simple
    - Compound
      - Intersecting compound equivalence
      - Cumulative compound equivalence
  - Hierarchical
    - Broader
    - Narrower
  - Associative

- “Exact” or “Inexact” applies to simple but not compound equivalence
Equivalence subdivisions with examples

- **Simple**
  - Laptop computers  EQ  Notebook computers

- **Compound**
  - Intersecting compound equivalence
    - Women executives  EQ  Women + Executives
  - Cumulative compound equivalence
    - Inland waterways  EQ  rivers | canals
Intersecting versus cumulative equivalence

Women executives  EQ  Women + Executives
Inland waterways  EQ  rivers | canals
And what about reciprocity?

- With simple equivalence (exact or inexact) and with hierarchical or associative mappings, two-way conversions are usually OK; but compound equivalence typically works in one direction only.

- ISO 25964 recommends that mappings from one vocabulary to another should be checked individually before they are used in the reverse direction.
cf. SKOS (Simple Knowledge Organization System) data model

Basic mapping “properties” (skos:mappingRelation):

- skos:closeMatch  (symmetric)
  - skos:exactMatch  (symmetric, transitive)
- skos:relatedMatch  (symmetric)
- skos:broadMatch   (inverse of narrowmatch)
- skos:narrowMatch   (inverse of broadmatch)

- No provision for compound mappings
- But much attention to reciprocity
In the real world, mapping perfection is elusive…

- Mapping projects are labour intensive, and often under-resourced
- Exact equivalence is all too rare
- Even when exact equivalence seems likely, it is often hard to be sure
- Often the vocabularies to be mapped are less than perfectly constructed
- Compound equivalence is needed commonly, but often unavailable
- Some systems allow only one mapping per concept
- While preparing mappings, you can’t make assumptions about capabilities of the search software
Practical Example 1

Vocabulary 1 has the following 3 concepts:
  Woods; Forests; Trees
Vocabulary 2 has only one approximate match:
  Woodlands

Mappings from Voc1 to Voc2:
  Woods   ~EQ Woodlands
  Forests ~EQ Woodlands
  Trees    RM Woodlands

Mappings from Voc2 to Voc1:
  Woodlands ~EQ Woods
              ~EQ Forests
              RM   Trees
**Practical Example 2**

<table>
<thead>
<tr>
<th>Thesaurus A</th>
<th>Thesaurus B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal products</td>
<td>Fats and oils</td>
</tr>
<tr>
<td>Edible fats and oils</td>
<td>Animal fats</td>
</tr>
<tr>
<td>Fats</td>
<td>Butter</td>
</tr>
<tr>
<td>Oils</td>
<td>Margarine</td>
</tr>
<tr>
<td>Vegetable products</td>
<td>Suet</td>
</tr>
<tr>
<td></td>
<td>Vegetable fats</td>
</tr>
<tr>
<td></td>
<td>Vegetable oils</td>
</tr>
<tr>
<td></td>
<td>Linseed oil</td>
</tr>
<tr>
<td></td>
<td>Olive oil</td>
</tr>
<tr>
<td></td>
<td>Sunflower oil</td>
</tr>
</tbody>
</table>
Mappings from A to B

Animal products  NM  Animal fats

Edible fats and oils  NM  Butter
  Margarine
  Olive oil
  Suet
  Sunflower oil

Fats  EQ  Animal fats | Vegetable fats

Oils  NM  Vegetable oils

Vegetable products  EQ  Vegetable fats | Vegetable oils
Mappings from B to A

Fats and oils  EQ  Fats | Oils
Animal fats  EQ  Fats + Animal products
Butter BM Animal products
   Edible fats and oils
   Fats
Margarine  BM Animal products
   Edible fats and oils
   Fats
Suet  BM Animal products
   Edible fats and oils
   Fats
Vegetable fats  EQ  Fats + Vegetable products
Vegetable oils  EQ  Oils + Vegetable products
Linseed oil  BM  Oils
   Vegetable products
Olive oil  BM  Edible fats and oils
   Oils
   Vegetable products
Sunflower oil  BM  Edible fats and oils
   Oils
   Vegetable products
a few conclusions

- If you prepare mappings from A to B, and rely on inverse mappings for the reverse direction, some concepts will not get mapped.
- Mappings that seem adequate for indexing applications may be insufficient for searching applications.
- Compound mappings do not work reciprocally, but could be used to suggest hierarchical mappings in the reverse direction.
- In systems which do not admit compound mappings, and/or which use only one mapping per concept, some very tough mapping choices have to be made.
So is SKOS right to rely on reciprocity in all its mappings?

- At the micro level, and in the absence of any compound mappings, reciprocity should work as per the SKOS guidance.
- But at the macro level, it is inadequate to prepare mappings in one direction and suppose this is enough for the reverse direction.
- There are too many uncertainties about the context in which mappings were prepared, and about the way search systems will exploit mappings and internal relationships within each vocabulary.
- ISO 25964 continues to recommend that if you want two-way mapping capability, the mappings should be checked individually before they are used in the reverse direction.
Want a copy of ISO 25964-2?

- A draft is due to appear in January 2011, "ISO DIS 25964-2", with the hope of attracting comments from potential users.
- The official way to get it is through your national standards body (e.g. BSI, DIN).
- Distribution policies vary from one country to another; last time round we found a way to make the draft available online free of charge and free of passwords, on the BSI site.
- Send me an email and I’ll alert you when the DIS is released. stella@lukehouse.org