Structural choices for KOS development, with Social Care Online as case study

Stella G Dextre Clarke
Information consultant
Summary

- Background; KOS lessons from the past
- Context of SCIE/SCO project
- Challenges to be addressed
- Structural options and choices
- Outcomes
A working IR system needs KOS embedded in a workflow jigsaw

Vocabulary management software

indexing/tagging software

search/browsing software
A few more things I’ve learned

- If you don’t make it easy for indexers, they won’t use a thesaurus properly;
- Plus they need training, incentives and quality control;
- As for searchers, it’s very difficult to get them to use a thesaurus at all;
- Nowadays, we must find ways to use KOS behind the scenes;
- Where KOS is exposed, it must be easy/appealing to navigate and apply;
- A taxonomy (browse tool) is usually more appealing than a thesaurus (search tool).
The context at SCIE in 2012, following a usability study

- Complete redevelopment of SCO database, including:
  - New search functionality
  - User interface
  - Content coverage review
- One component to be a review and redevelopment of existing SCO subject taxonomy
- NB in view of funding cutbacks, SCIE may need to adjust its business model
Taxonomy review terms of reference (abbreviated)

- Review the current structure, especially the top terms;
- Advise on a polyhierarchical approach;
- Recommend a structural model that allows for regular updating, in view of frequent terminological and structural changes in the sector;
- Advise on the use of compound terms, pre- and post-coordinated indexing;
- Improve accuracy of hierarchical relationships, so that search explosion gives better results;
- Review coverage of current terms and highlight gaps in coverage.
Terms of reference (structural aspects emphasis)

- Review the current **structure**, especially the top terms;
- Advise on a **polyhierarchical** approach;
- Recommend a **structural model** that allows for regular updating, in view of frequent terminological and structural changes in the sector;
- Advise on the use of compound terms, pre- and post-coordinated indexing;
- Improve accuracy of **hierarchical relationships**, so that search explosion gives better results;
- Review coverage of current terms and highlight gaps in coverage.
Interface top level in 2012

Welcome ...

to the UK’s largest database of information and research on all aspects of social care and social work. Updated daily resources include legislation, government documents, practice and guidance, systematic reviews, research briefings, reports, journal articles and websites. Every resource listed includes an abstract. Links to full text are also included where available. Social Care Online is a particularly useful resource for staff, students and researchers working in social work and social care.

Simple search

Enter a single word OR a phrase

Click here for help with simple searching.
Click here for intermediate search or advanced search.

My Social Care Online services

Services include email alerts to the latest social care publications and resources, publicising events on Social Care Online, and creating your own reading lists.

Log-in or sign-up to find out more.

SCIE Athens service

The SCIE Athens service provides free access to a selective range of social care and health journals and databases.

An Athens account increases the amount of the full-text content available through Social Care Online.

Find out more about SCIE Athens.

Eligibility for SCIE Athens is restricted to local authority social care staff.

Visit the SCIE website to find out if you are eligible and to sign-up for SCIE Athens.

Browse social care topics

benefits and personal finance
criminal justice, law and rights
education, training and employment
families, children and young people
government and social policy
health and health care
housing and environment
living and life events
management and organisational development
mental health and mental health care
people, groups and communities
physical and learning disabilities
research and evaluation
social care services
social work and social workers

Find topic

Click here for help with browsing topics.
Taxonomy expansion was easy - and popular
## Parameters of existing taxonomy

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred terms</td>
<td>1099</td>
</tr>
<tr>
<td>Non-preferred terms</td>
<td>680</td>
</tr>
<tr>
<td>Top terms</td>
<td>15</td>
</tr>
<tr>
<td>Associative relationships</td>
<td>418</td>
</tr>
<tr>
<td>Scope notes</td>
<td>246</td>
</tr>
<tr>
<td>History notes</td>
<td>0</td>
</tr>
<tr>
<td>Node labels</td>
<td>0</td>
</tr>
<tr>
<td>Polyhierarchy?</td>
<td>No</td>
</tr>
</tbody>
</table>
Existing approach had evident strengths, but underlying weaknesses:

- Top level headings a mix of facets and subject areas, with potential overlaps
- Hard to predict where to browse to find a given concept
- Lack of hierarchical rigour makes search explosion fallible
- Unclear if the KOS is really a taxonomy or a thesaurus
- Unclear whether you are browsing a resource collection or a vocabulary
Top level headings a mix of facets and subject areas

- benefits and personal finance
- criminal justice, law and rights
- education, training and employment
- families, children and young people
- government and social policy
- health and health care
- housing and environment
- living and life events
- management and organisational development
- mental health and mental health care
- people, groups and communities
- physical and learning disabilities
- research and evaluation
- social care services
- social work and social workers
Hierarchical content hard to predict; explosion results weird

families, children and young people
  . babies
    .. premature babies
  . child care
  . child day care
    .. after school care
  . child minding
  . nurseries
  . playgroups
  . child development
    .. adolescence
  . child support
  . children
    .. adopted children
    .. boys
    .. children in need
    . emotionally disturbed children
  .. foster children
    .. girls
    .. hyperactive children
  . looked after children
  . pre-school children
  . school children
  . vulnerable children
  . contact
  . child custody
    ... joint custody
  . families
    .. birth families
    .. single parent families
    .. step families
  . family members
    .. adoptive parents
    .. birth parents
    . fathers
    . mothers
    . non-resident parents
    . parents with learning disabilities
    . separated parents
    . teenage parents
  . relatives
    .. siblings
    .. triplets
    .. twins
  . family relations
    . family reunification
    . parent-child relations
    . step relationships
  . parenting
    . discipline
    . parental responsibility
    . parental rights
    . parental role
  . pregnancy
    . abortion
    . multiple pregnancy
    . surrogacy
    . teenage pregnancy
  . unborn children
  . runaways
  . wardship
  . young adults
  . young people
Examples of relationship problems in existing taxonomy

psychiatric day centres
BT psychiatric care
RT day centres

prostitution
UF prostitutes
    sex workers
BT crime

diabetes
UF diabetics
BT diseases
Are you browsing a collection or a vocabulary? Thesaurus or taxonomy?
Archetypal search interface
Dyed-in-the-wool browsing
Combination of searching and browsing; taxonomy hidden

650,030 items for:
chips

1. **Special Collection** **DENMARK POTATO CHIPS**
   60% - Industry overviews: Report of DATAMONITOR 01/01/2003
   MarkIntel Market Research (full reports): Available at NorthernLight.com

2. **Special Collection** **Natural chips**
   59% - Articles & General info: Lamb's Natural Chips are another way to fill your side dish needs with quality Lamb Weston potato products. These delicious potato chips allow you ...
   09/01/2002
   Food Processing (magazine): Available at NorthernLight.com

3. **Special Collection** **Blue Chip Value Fund Declares Year-End Distribution**
   58% - Press releases: DENVER, Dec 16, 2002 (BUSINESS WIRE) -- The Directors of Blue Chip Value Fund, Inc, have declared a distribution of $0.12 per share. Of this ...
   12/16/2002
   Business Wire (newswire): Available at NorthernLight.com

4. **Special Collection** **Blue Chip Blues**
   58% - Articles & General info: Show: CNNfn NEWS Date: 12/12/02 12/12/2002
   CNNfn Newslink (transcripts): Available at NorthernLight.com

5. **Special Collection** **CHIP REIT**
   57% - Company information: Report of RBC CAPITAL MARKETS (CANADA)
   11/15/2002
Some more little things I’ve learned

- People often like to search/browse/filter in combination.
- It’s good to see some sample hits as you go along.
- Faceted search comes intuitively only if the facets are mutually exclusive.
- A KOS optimised for browsing is not usually optimal for searching, and vice versa.
Here you can’t see hits as you browse; you can’t use categories to disambiguate
Enough of the challenges; what to do?

- Decide what functions the vocabulary needs to support:
  - searching?
  - browsing?
  - filtering?
  - a combination?

- NB Starting with a clean slate is not an option; the existing terms have been used to index ~150,000 items
Structural options discussed

A. Restructure the taxonomy as a faceted thesaurus, using strict rules of hierarchy

B. Split the taxonomy into two vocabularies, comprising index/search terms and broad browsing categories respectively. This option can be further subdivided into two, depending on whether:
   1. the subject/browsing categories are just a short unstructured list of filters, or
   2. the subject categories form a structured taxonomy, 2-3 levels deep.

C. Hybrid taxonomy, with the top levels mostly comprising broad subject terms and loose rules of hierarchy, but stricter rules are followed where reliable search explosion is needed.
Option C: Hybrid taxonomy – the least change option

- This would deliver thesaurus power for search terms at the lower level, with explosion working properly, while still supporting browsing at the top 2 levels
- Customer loyalty maintained
- Need for software change minimised
- But – doesn’t resolve all the observed problems. It’s a fudge.
Option B1: Split into two – a proper thesaurus + a set of subject filters

- Following ISO25964-1 rules leads to a stronger search vocabulary;
- Search explosion works properly;
- More RTs to support lateral navigation;
- Top terms could be 100+ and it would not matter;
- Disambiguation enabled by subject filters, e.g:

<table>
<thead>
<tr>
<th>adult services</th>
<th>mental health</th>
</tr>
</thead>
<tbody>
<tr>
<td>children’s/family services</td>
<td>disability</td>
</tr>
<tr>
<td>older people</td>
<td>direct care</td>
</tr>
</tbody>
</table>
Option B2: As B1, but filters expanded to 2-3 levels deep

- Same advantages as B1, plus
- Affords opportunity for browsing database.
- Could match some categories to key user groups.
- Disambiguation opportunities enhanced.
- But more work is involved – user testing and software development as well as taxonomy construction.
Option A: Fully faceted thesaurus

- A Rolls-Royce thesaurus, with ISO 25964-1 rules followed and the structural problems fixed

- Main difficulties are:
  - Learning curve for users (and staff)
  - Lack of tools for construction
  - Lack of tools for implementation, including export/import

- NB you don’t need all the power of faceted classification if you don’t have books to shelve in linear order.
Option B2 was chosen

- In summary - Split the old taxonomy into two vocabularies: a thesaurus of index/search terms and a taxonomy of broad browsing categories structured to 2-3 levels.
- The new taxonomy should respond to an earlier consultant’s call for effective filter categories:
  “The ability to filter out information specific to children’s or adults’ services would have significantly enhanced SCO’s effectiveness for the majority of user journeys conducted in this review”
Work began – based on painstaking scrutiny of how index terms had been used.
# Vocabulary parameters before and after

<table>
<thead>
<tr>
<th>Category</th>
<th>Existing taxonomy</th>
<th>New thesaurus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred terms</td>
<td>1099</td>
<td>1159</td>
</tr>
<tr>
<td>Non-preferred terms</td>
<td>680</td>
<td>1766</td>
</tr>
<tr>
<td>Top terms</td>
<td>15</td>
<td>49</td>
</tr>
<tr>
<td>Associative relationships</td>
<td>418</td>
<td>3116</td>
</tr>
<tr>
<td>Scope notes</td>
<td>246</td>
<td>305</td>
</tr>
<tr>
<td>History notes</td>
<td>0</td>
<td>124</td>
</tr>
<tr>
<td>Node labels</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Polyhierarchy?</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Next steps

- Develop small new browsing taxonomy
- Back-file clean-up
- Resolve remaining queries
- Integrate thesaurus with search functionality and interfaces
- Integrate thesaurus with indexing module
- Maintain thesaurus
- Training strategy
Don’t forget the KOS needs embedding in a workflow jigsaw

Vocabulary management software

Indexing/tagging software

Search/browsing software
Potential benefits of vocabulary implementation

- Removal of user confusion about what is being browsed - the vocabulary per se, or the database content;
- Search precision increased, as a result of improved specificity and resolution of ambiguous search terms;
- Search recall improved, as synonyms and relationships are exploited in simple search;
- Selective search expansion available on demand;
- “Did you mean..” choices informed by near-matches from the thesaurus;
- Search refinement easier, when taxonomy filters remove irrelevant noise;
More potential benefits of vocabulary implementation

- Results ranking/clustering, exploiting thesaurus terms/structure;
- Clarification of specialised terminology available via display of enhanced scope notes;
- Indexing quality improved as a result of thesaurus navigation tools.
- Formulation of complex search statements easier for advanced users with access to thesaurus navigation interface;
- Search explosion available for experienced users;
- Thesaurus navigation features can be presented as an extra benefit to subscribers
The new site – a work in progress?

Social Care Online
Social Care Online is the UK’s largest database of information and research on all aspects of social care and social work.

About Social Care Online
Social Care Online is produced by the Social Care Institute for Excellence (SCIE) and is the UK’s largest database of information on all aspects of social care and social work.

- What does Social Care Online cover?
- Which journals are indexed on Social Care Online?
- Find out more about Social Care Online

Key features
- Fast Google-type standard search, with the ability to filter searches
- An advanced search option offering saved searches and email alerts for registered users
- Informative abstracts with links to the full text of reports, and other publications if these are freely available on the web

Perform a standard search (above) or use the advanced search

Using Social Care Online
Whatever your role in social care, you need to keep up to date with the latest information, knowledge and good practice. Social Care Online provides easy access to a wide range of high quality information, research and hard-to-find resources which have been evaluated by social care information specialists. The database is updated daily and is also free to use!

Get help using Social Care Online

Social Care Online
- Home
- Advanced Search
- Using our data
- About Social Care Online
- Register/agm-in

Using the site
- Help using the website
- Accessibility
- Terms and conditions
- Cookies

Connect with us
- Facebook
- Twitter

© Social Care Institute for Excellence. All rights reserved [1.1.7.0]