Signposting the crossroads: terminology web services and classification-based interoperability

Gordon Dunsire and Dennis Nicholson

University of Strathclyde

Presented at Classification at a crossroads: multiple directions to usability, International UDC Seminar 2009, The Hague, Netherlands, 29-30 Oct 2009

Overview

✧ HILT project, architecture and database

✧ Embedding HILT in end-user services

✧ Developing a distributed approach

✧ Potential role of UDC

HILT project

- ✧ High-Level Thesaurus, funded by UK's Joint Information Systems Committee (JISC) 2000-2009, in 4 phases
- ✧ Investigated issues in finding stuff indexed by different subject schemes
  - ✧ Primarily monolingual (i.e. Anglophone)
- ✧ Developed some solutions
  - ✧ Use of classification-based switching spine
    - ✧ Dewey Decimal Classification (DDC)
  - ✧ Subject-based identification of collections and matching terms in the subject schemes used
  - ✧ Machine-to-machine terminology services

HILT architecture

Database of terminologies mapped to DDC notations

⇕

Application Programming Interface (API) via SOAP server

⇕

M2M access via SRU/W server

⇕

Client applications

HILT API

✧ Input DDC notation > Output matching collection-level metadata, including item-level subject scheme

✧ Input subject term > Output mapped DDC notations and captions

✧ Input DDC notation > Output mapped terms

✧ Input term and terminology > Output matching and related terms

✧ Input term > Output terms with similar spellings

✧ Input term > Output definitions and descriptions

HILT terminologies database

✧ Hub-and-spoke approach to mapping
  ✧ DDC notation as hub (spine, switching language)
  ✧ Terms and non-DDC notations as spoke

✧ Avoids direct term-term mappings
  ✧ And the resulting combinatorial explosion as the number of terminologies increases

✧ Notation as hub is language-independent and provides accommodation for new terms

✧ But indirect mappings (term-term) may exhibit semantic misalignment (drift)

Embedding HILT in end-user services

- ✦ Extension to final phase of HILT
  - ✦ Pilot embedding in three operational services
    - ✦ User-generated subject metadata (The Depot)
    - ✦ Multi-scheme digital library (Intute)
    - ✦ Collection landscaping (SCONE)
- ✦ API functions combined as higher-level applications
  - ✦ Spell-checking
  - ✦ Disambiguation of user term
  - ✦ Mapping user term to controlled terms
  - ✦ Suggesting related terms

HILT embedded in SCONE:
http://scone.strath.ac.uk/Service/SconeServiceHilt/Hiltdemo.cfm

Scottish Collections Network

About | Accessibility | Using SCONE | Glossary | Disclaimer
Home | Titles | Names | Subjects | Locations | Landscapes | Help
DDC subject term results for your search term.

Search term	teeth

Enter a search term above, then press Go

All DDC subject terms will be returned for the search term you entered. You can also search for Welsh or German terms (where available)

Example: Enter "teeth" to search for subject terms relating to "teeth". You can then click on the specific subject term that you choose is most relevant to then display the collections available for that specific term.

To erase the search term press Clear

Search Suggestions:

Natural sciences and mathematics > Life sciences Biology > Internal biological processes and structures > Specific physiological systems in animals, regional histology and physiology in animals > Digestive system > Mouth and esophagus > Teeth

---

Natural sciences and mathematics > Life sciences Biology > Natural history of specific kinds of organisms > Plants and animals > Animals (Zoology) > Specific taxonomic groups of animals > Mammalia (Mammals) > Eutheria (Placental mammals) > Hominidae Homo sapiens > Anthropometry > **Teeth**

Technology (Applied sciences) > Medicine and health > Human anatomy, cytology, histology > Gross anatomy > Digestive tract organs > Mouth > **Teeth**

Technology (Applied sciences) > Medicine and health > Human physiology > Specific functions, systems, organs > Digestion > Mouth and esophagus > **Teeth**

Technology (Applied sciences) > Medicine and health > Diseases > Specific diseases > Other diseases > Tumors and miscellaneous communicable diseases > Cancers > Cancers of other organs and of regions > Digestive system diseases--humans--cancer--medicine, . . . > Mouth cancer--medicine, . . . > **Tooth diseases--humans--cancer**

Technology (Applied sciences) > Medicine and health > Miscellaneous branches of medicine Surgery > Dentistry > Dental diseases > **Tooth diseases--humans--therapy**

Titles
(35 in list)

Selected Collection titles

Aberdeen University Medical Library collection
Aberdeen University Medical Library SWOP - WHO collection
Advocates' Library SWOP - WHO collection
Argyll and Bute Hospital Medical Library collection
Douglas papers
Dundee University Ninewells Medical Library collection
Edinburgh City Libraries. Central Library SWOP - WHO collection
Edinburgh University Library. Main Library SWOP - WHO collection
FL collection
Glasgow Royal Infirmary Library and eLearning Centre collection

To display details of a specific collection, select it from the list, then | Go |

To search for items in the listed collections using the Cooperative Information Retrieval Network for Scotland,
| Search CAIRNS |

To search for items in the listed collections using other online catalogues, | Search OPACs |

Collection types
- ☑ Archive
- ☑ Museum
- ☑ Library
- ☑ Internet

To change the type of collection listed, select one or more collection types, then | Change list |

Developing a distributed approach

- ✧ Scaling remains a factor, even in hub-and-spoke architectures
  - ✧ Thousands of terms from hundreds of vocabularies
- ✧ Machine mappings using statistical association can help
  - ✧ OCLC's WebDewey mapping of DDC to Library of Congress Subject Headings (LCSH) uses associations in WorldCat
- ✧ User-generated metadata may also help
- ✧ Semantic Web and linked data

Semantic Web

- ✦ LCSH and Rameau (French) available as linked data in Simple Knowledge Organization System (SKOS) format
  - ✦ DDC top 3 levels (000-999) available as open linked data
    - ✦ Representation of the whole scheme in development
- ✦ Library of Congress Name Authority File and Virtual International Authority File
- ✦ Representation of metadata formats will allow parsing of instance records as linked data
  - ✦ Critical mass for statistical mapping

Distributed architecture

Distributed linked data for subject terminologies

⇕

SPARQL + distributed SPARQL servers

⇕

Client applications

Potential role of UDC

✧ Machine-readable version available ✓

✧ Covers all subject areas ✓

✧ In widespread use ✓

✧ Language translations ✓

✧ Mappings to other schemes exist ✓

✧ RDF/linked data representation ?

✧ Open availability ?

✧ Critical mass of catalogue records using UDC and other subject schemes ?

Thank you

- ✧ g.dunsire@strath.ac.uk