Terminology services and the DDC: the High-Level Thesaurus and beyond

Presented to the symposium Dewey goes Europe: on the use and development of the Dewey Decimal Classification (DDC) in European libraries

28 April 2009, Austrian National Library, Vienna

Gordon Dunsire

Overview

- ✧ Terminology services
- ✧ HILT
- ✧ To the Semantic Web and beyond

Terminology services

- ✧ "Web services involving various types of knowledge organization resources [vocabularies], including authority files, subject heading systems, thesauri, Web taxonomies, and classification schemes"
  - ✧ Vizine-Goetz and others (OCLC), 2004
  - ✧ Example: Mappings from a term in one vocabulary to one or more terms in another vocabulary

HILT: High-Level Thesaurus project

- ✧ Started in 2000. Now in 4th phase, embedding functionality in pilot services
    - ✧ Funded by Joint Information Systems Committee, and supported by OCLC
- ✧ To provide subject interoperability in a multi-scheme environment via inter-scheme mapping
    - ✧ Ideally by identifying a generic approach that can be developed through distributed collaborative action …

Role of DDC in HILT

- ✦ **DDC used as a semantic translator between different Anglophone subject terminologies**
  - ✦ ☺ Availability of DDC mappings to captions, relative index and LCSH
- ✦ **DDC (partially) mapped to other terminologies**
  - ✦ AAT, Unesco, MeSH, etc. + some non-English
- ✦ **DDC also used to classify collections**
  - ✦ "Subject landscaping" gives best places (catalogues) to search for items
  - ✦ Drill-up: (Item > Collection) = (Specific > General) = Truncation of DDC notation

HILT applications: Disambiguation

- ✧ **User-supplied subject term may be ambiguous**
  - ✧ E.g. "tree" – genealogy or forestry?
- ✧ **So, display all subject headings or captions containing the term, along with context**
  - ✧ Hierarchy of "broader" headings
    - ✧ Driven by DDC notation
  - ✧ Non-HILT possibilities include definitions, scope notes, etc.
- ✧ **User chooses heading/caption**
  - ✧ System knows mapped DDC notation

HILT applications: Spell-checking

✧ User-supplied subject term may have incorrect spelling, or variant spelling

✧ E.g. "colour" vs "color"

✧ So, display terms from all HILT vocabularies which pattern-match the user term

✧ User is alerted if term is erroneous, or variant spellings exist, or term is not present in any vocabulary

HILT applications: Item-level retrieval

- ✧ System knows what subject heading or classification scheme is indexed in a specific catalogue or other item-level finding-aid
- ✧ User-supplied subject term is matched to semantically-equivalent term in the scheme
- ✧ Matched term is used to search finding-aid
  - ✧ Manually, by user
  - ✧ Transparently by service, using machine-to-machine interface for finding-aid, if available

HILT applications: Landscaping

- ✧ "Landscape" is a set of collections and associated finding-aids which are strong in a specific subject area
- ✧ Classify collection-level descriptions with DDC
    - ✧ More than one notation may be required
- ✧ Use implicit hierarchy of decimal notation to identify collections containing a user-supplied term
    - ✧ User term: teeth = 611.314
    - ✧ 611 = Anatomy museum collection

Putting it all together

- ✧ User supplies term
- ✧ Term is disambiguated (and spell-checked)
- ✧ Collections with strength in subject area of term are identified
- ✧ Collection finding-aids are searched with semantically-matched terms appropriate to their specific subject scheme
- ✧ User gets item-level metadata from multiple collections

Example collections service: SCONE

- ✦ **Scottish Collections Network**
    - ✦ Metadata for collections located in Scotland
    - ✦ Archives, libraries, museums, etc.
- ✦ **Database drives a set of services collectively known as "Scotland's Information Landscape"**
- ✦ **Collections with specific subject focus are classified with DDC**
    - ✦ Multiple notations used if necessary
- ✦ **DDC summary captions used to display a hierarchical, browsable tag cloud**

File   Edit   View   History   Bookmarks   Tools   Help

Scotland's Information

About

Find it

Where

Libraries

Archives

Museums

What

Topics

People

Organisations

How

Languages

Special

Landscapes

Tag cloud of DDC level 1 topics

Ⓐ Ⓛ Ⓜ

* Religion *   * Science *   * Technology *

* Computer science, information & general works *

* History & geography *   * Social sciences *

* Arts & recreation *   * Literature *   * Language *   * Philosophy & psychology *

Use the link on an element in the tag cloud to restrict the scope of your search.

Display tags in alphabetic order

The DDC (Dewey Decimal Classification) divides subjects into 10 topics at level 1. Each level 1 topic has 10 subdivision topics at level 2, and each subdivision is in turn divided into 10 topics at level 3.
All copyright rights in the Dewey Decimal Classification system are owned by OCLC. Dewey, Dewey Decimal Classification, DDC and WebDewey are registered trademarks of OCLC.

Done

File   Edit   View   History   Bookmarks   Tools   Help

Scotland's Information

About

Find it

Where
- Libraries
- Archives
- Museums

What
- Topics
- People
- Organisations

How
- Languages

Special
- Landscapes

Tag cloud of DDC level 2 topics for Arts & recreation

(A) (L) (M)

* Sculpture, ceramics & metalwork *   * Graphic arts *   * Architecture *

* Sports, games & entertainment *

* Drawing & decorative arts *   * Arts *   * Music *

* Photography & computer art *   * Painting *   * Landscaping & area planning *

Use the link on an element in the tag cloud to restrict the scope of your search.

Display all collections associated with this tag cloud.

Display tags in alphabetic order

The DDC (Dewey Decimal Classification) divides subjects into 10 topics at level 1. Each level 1 topic has 10 subdivision topics at level 2, and each subdivision is in turn divided into 10 topics at level 3.
All copyright rights in the Dewey Decimal Classification system are owned by OCLC. Dewey, Dewey Decimal Classification, DDC and WebDewey are registered trademarks of OCLC.

Done

Scotland's Information

About

Find it

Where

Libraries

Archives

Museums

What

Topics

People

Organisations

How

Languages

Special

Landscapes

Tag cloud of DDC level 3 topics for Sports, games & entertainment

(A) (L) (M)

* Recreational & performing arts *   * Stage presentations *   * Aquatic & air sports *

* Athletic & outdoor sports & games *

* Fishing, hunting & shooting *   * Indoor games of skill *   * Public performances *

Use the link on an element in the tag cloud to show a map of locations

Display all collections associated with this tag cloud.

Display tags in alphabetic order

The DDC (Dewey Decimal Classification) divides subjects into 10 topics at level 1. Each level 1 topic has 10 subdivision topics at level 2, and each subdivision is in turn divided into 10 topics at level 3.
All copyright rights in the Dewey Decimal Classification system are owned by OCLC. Dewey, Dewey Decimal Classification, DDC and WebDewey are registered trademarks of OCLC.

Done

File   Edit   View   History   Bookmarks   Tools   Help

Scotland's Information

✳ **About**

Information collections about
Athletic & outdoor sports & games

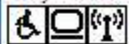22 listed

Humble collection
♿ 🖥 📡 Where? What?

Jim Clark Room collection
Where? What?

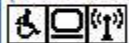Lloyd collection
♿ 🖥 📡 Where? What?

Luffness New Golf Club
(Acc.11692) collection
♿ 🖥 📡 Where? What?

Luffness New Golf Club
(Acc.12628) collection
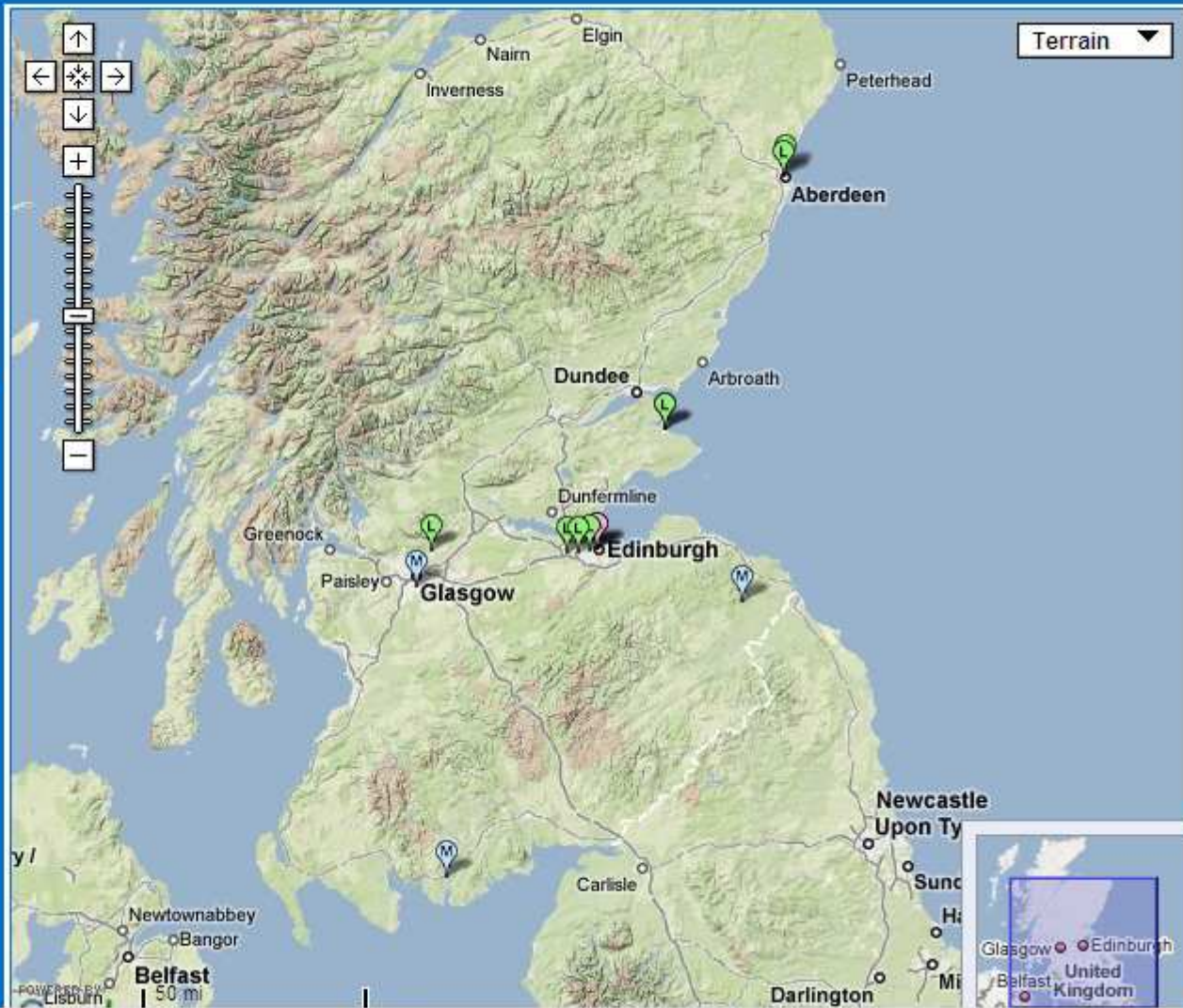♿ 🖥 📡 Where? What?

North Berwick Golf Club collection
♿ 🖥 📡 Where? What?

Royal Caledonian Curling Club
Library collection
Where? What?

Terrain ▼

Elgin
Nairn
Inverness
Peterhead
Aberdeen
Dundee
Arbroath
Dunfermline
Greenock
Paisley
Glasgow
Edinburgh
Newcastle Upon Ty
Carlisle
Sund
Newtownabbey
Bangor
Belfast
50 mi
Lisburn
Darlington
Mi

Glasgow Edinburgh
Belfast United Kingdom

Done

Applying HILT to SCONE

- **Replace hard-wired DDC summary captions (tags) with captions from HILT**
  - Problem: HILT assumes minimum 3-digit notation, and currently cannot cope with 3xx, 33x, etc.
- **Add disambiguation and landscaping facilities**
- **Add suggested terms for searching specific finding-aids**
  - SCONE already stores metadata about subject schemes
- **Intute service (project partner) pilot available**

HILT embedded in Intute: http://www.intute.ac.uk/search_hilt.html

Search Results

Search [tree] in [All subjects ▾] Go

Advanced search Subject A-Z New resources Help

Search term(s)	**tree** [subject(s): **All**]
Related terms [max 10]	genealogy broadleaved forests coniferous forests countryside conservation forest management forestry mixed forests orchards soft fruit sylviculture
HILT spell	trek tres free cree true
Intute catalogue results	**533** [By group: **A** 90 **H** 350 **S** 73 **So** 20 (What's this?)]
Harvester results	**View results** (What's this?)

Filter by resource type: [All ▾] Go

S Sciences **A** Arts **So** Social **H** Health

Now showing: **1 - 25** of 533 records
Page: **1** 2 3 4 5 ... 22 >> Next   Order by **Relevance** | Alphabetical

Tree-Ring Data Search   **S** More details Save record

The **Tree**-Ring Data Search gives access to the International **Tree** Ring Data Bank, a palaeoclimatic database maintained by the NOAA Paleoclimatology Program and World Data Center for

Beyond HILT

- ✧ HILT approach too expensive to scale across all subject schemes
  - ✧ Hence "High level"
- ✧ Might be complemented by other methods
  - ✧ One-one vocabulary mappings
  - ✧ Associative clusters and mappings
    - ✧ Similar to DDC-LCSH in WebDewey
- ✧ Wider context of the Semantic Web
  - ✧ SKOS vocabularies
  - ✧ RDFS/OWL ontologies and schemas

Behind HILT

✧ DDC licensing issues are a barrier to full exploitation of HILT

  ✧ And other terminology services

    ✧ Including OCLC TS pilot?

✧ Tension between Semantic Web "open-ness" and OCLC/DDC cost recovery

✧ Creative Commons Attribution-NonCommercial-ShareAlike license would help

  ✧ Or something similar

In front of HILT

- ✧ HILT designed as machine-to-machine service
  - ✧ HILT "end-users" are other online services
  - ✧ OCLC Terminology Services pilot also intended for technical consumption
- ✧ Emerging need for terminology services for non-professionals
  - ✧ E.g. Simple drop-down term lists; simple thesaurus browse; etc.
  - ✧ Generic "widgets", plug-ins, etc.?

Immediately in front of HILT

- ✧ DDC translations!
- ✧ HILT infrastructure is independent of language
  - ✧ Already has non-English (e.g. Welsh) terms
- ✧ Should be technically easy to incorporate translations
  - ✧ Just another DDC-mapped vocabulary
- ✧ Full depth not required
  - ✧ "High" level

Thank you!

- ✧ G.dunsire@strath.ac.uk
- ✧ HILT
  - ✧ http://hilt.cdlr.strath.ac.uk/
- ✧ Scotland's Information
  - ✧ http://www.scotlandsinformation.com/
- ✧ OCLC Terminology Services pilot
  - ✧ http://tspilot.oclc.org/resources/