Revision and extension of thesaurus standards

Stella G Dextre Clarke

Alan Gilchrist

Leonard Will

Existing thesaurus standards

- ISO 2788-1986 Guidelines for the establishment and development of monolingual thesauri
    - = BS 5723:1987
- ISO 5964-1985 Guidelines for the establishment and development of multilingual thesauri
    - = BS 6723:1985
- ANSI/NISO Z39.19-1993 Guidelines for the construction, format and management of monolingual thesauri

What needs updating?

- Printed *versus* electronic application
- Guidance on management software
- Interoperability:
    - Mapping between thesauri and other types of vocabulary
    - Formats/protocols for data exchange with downstream applications
- Applicability to end-user applications, not just those for information professionals

Procedural aspects

- Constituting a working group for an international standard takes a while
- It seemed easier to work first on the British Standards; then offer these up to the international community
- The BS committee must be UK-based, but invites international input wherever possible without offending BSI/ISO

Procedure in detail

- Small UK Working Group of unpaid volunteers submits drafts to BSI committee IDT/2/2
- Comments are invited informally from everyone interested
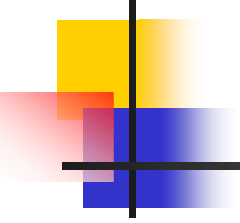- Progress is very slow... but worthwhile, we hope

Outline of new standard

- BS 8723: Structured vocabularies for information retrieval – Guide
- Part 1: General
- Part 2: Thesauri
- Part 3: Vocabularies other than thesauri
- Part 4: Interoperability between vocabularies
- Part 5: Interoperability with applications

Progress to date

- Drafts of Parts 1 and 2 soon to be issued for public comment
- Draft chapters of Part 3 exist, including Classification schemes, Taxonomies, Subject heading lists, Ontologies, Search thesauri
- Draft of Part 4 well advanced
- Part 5 to follow the rest

New features in Part 2 (monolingual thesauri)

- Clearer guidance on applying facet analysis to thesauri
- Some changes to the 'rules' for compound terms
- More guidance on managing thesaurus development and maintenance
- Functional specification for software to manage thesauri
- General overhaul in all areas

Part 4: Interoperability between vocabularies

- Huge demand for accessing information that has been indexed with another language and/or vocabulary. The buzzword is 'Mapping'. The Semantic Web is just one application.

- Part 4 to include multilingual thesauri as a special case of mapping between vocabularies

- But how do you map between precoordinate and postcoordinate schemes?

- How do you provide for mapping when notations are not all enumerated, but built up by rules for synthesis?

Part 3 chapters

- Classification schemes
- Subject heading lists
- Taxonomies
- Ontologies
- Semantic nets (?)
- Search thesauri

Issues for Part 3

- How much guidance is needed on how to build other sorts of vocabulary?
- Should we describe the idiosyncrasies of existing schemes, even where we judge there is a 'better' way?
- To provide a basis for Part 4, Part 3 should pick out the characteristics of different vocabulary types that govern when and how you can map them. But some of the observable characteristics might not be what we'd recommend. What to do?

Part 5: Interoperability with applications

- Vocabularies must work with
    - Search engines
    - Content Management Systems
    - Web publishing software, etc.
- Build on existing formats and protocols for data exchange
- e.g. Z39.50 and Zthes, XML schema? DTD? MARC? SKOS Core Schema? Topic Map? ADL gazetteer protocol? Anything else?

Want to get involved?

- We very much need wider input, so please get in touch!
- Email us at
  SDClarke@LukeHouse.demon.co.uk
  cura@fastnet.co.uk
  L.Will@willpowerinfo.co.uk