Knowledge organisation: classifications and thesaurus systems

Introduction
Knowledge organisation: introduction

- To organise knowledge / documents / books / reports / information / data / records / things / items / materials for more efficient storage and retrieval, some related, similar tools / systems / methods / approaches are used.
- Often but not yet always, this process is assisted by a computer system.
- Good systems are expanded and updated when the need arises.
- The organization system applied should ideally be clearly and immediately visible or even searchable on computer, by the user of the materials.
Knowledge organisation: some tools

- Various related tools / systems / methods / approaches are available:
  - Subject-related metadata
  - Classifications
  - Controlled list of selected keywords = authority files = controlled vocabularies
  - Taxonomies
  - Thesauri
  - Faceted classifications
  - Ontologies; topic maps
  - …
Knowledge organisation: classifications and thesaurus systems

Classifications
Classification systems: introduction

- Classification systems present the subjects in a logical order, usually going from the more general to the more specific.
Classification systems: examples of universal systems

• Universal means here: covering all subjects
• Not just one but several competing systems exist.

Examples

» *Universal Decimal Classification* = *UDC*
  used mainly outside U.S.A.

» *Dewey Decimal Classification* = *DDC*
  used mainly in U.S.A.

» *Library of Congress Classification*
  used mainly in U.S.A.

» ...
Knowledge organisation: classifications and thesaurus systems

Thesaurus systems
Thesaurus: description

- Thesaurus (contents) =
  » system to control a vocabulary
    (= words and phrases + their relations)
  »+ the contents of this vocabulary

- Thesaurus program =
  program to create, manage, modify and/or search a thesaurus using a computer
Which applications do you see for a thesaurus?
Thesaurus applications related to information searching (1)

- For producers of a database:
  To find/choose index terms to add these to items in a database, when terms are taken from a controlled vocabulary to increase precision and recall in the searches by users of the database.
Thesaurus related to information searching (2)
Thesaurus applications related to information searching (3)

• For users (!) of a database:
  When the database to be searched is produced with added descriptors (words and terms) that are taken from a controlled list of approved, selected words and terms, then the searcher can use some printed or computer-based system first, to find more and ‘correct’ suitable words and terms that belong to that controlled list of descriptors; then, the searcher can use these descriptors (and only these words or terms) in a database query.
Thesaurus applications related to information searching (4)

• For users (!) of a database:
  When the database to be searched is NOT produced with added descriptors (words and terms) that are taken from a controlled list of words and terms, then the searcher can use one or several thesaurus systems first, to find more words and terms and more suitable words and terms; then the searcher can use these found words and terms to formulate a query for that database (to increase recall and precision).
General thesaurus systems that cover all subjects

- General systems
- Universal systems
- Covering all subjects
- Broad and shallow systems
- Horizontal systems
General thesaurus systems: examples

- *Library of Congress Subject Headings (LCSH)*
- thesaurus system built into word processing software
- thesaurus system that runs on a pc (independent of Internet)
  see for instance [http://www.wordweb.co.uk/free/](http://www.wordweb.co.uk/free/)
General thesaurus systems on the WWW: examples

- thesaurus systems that can be used free of charge through the WWW
  - http://www.answers.com/library/Thesaurus
  - http://thesaurus.plumbdesign.com/
    up to early 2005 available free of charge, based on WordNet:
  - http://wordnet.princeton.edu/
General thesaurus system on pc: screenshot
General thesaurus systems: comments

- An ideal, complete, general thesaurus that covers all subjects does not exist.
!! Task - Assignment - Exercise !!

Try to find suitable search terms to retrieve documents on “pollution” from a database on marine science, by using for instance the thesaurus included in the program for word processing that you use.
Try to find suitable search terms to retrieve documents related to the concept “sea” from a database on marine science, by using for instance the thesaurus included in the program for word processing that you use.
Thesaurus systems focused on a particular subject

- Focused on a particular subject domain = narrow and deep, vertical systems
Thesaurus systems focused on a particular subject: examples

- **ERIC**: education, information science...
- **Psychological Abstracts / PsycInfo**
- **Sociological Abstracts / SocioFile**
- **INSPEC**: physics, electronics, information technology
- **Medline (the Medical Subject Headings = MeSH)**
Thesaurus systems focused on a particular subject: examples

• the *Aquatic Sciences and Fisheries Information System*

• Various thesaurus systems for art and architecture can be found online:
  http://www.getty.edu/research/tools/vocabulary/
Give an example of a horizontal thesaurus for the whole English language.
Suppose that you use a search system which is NOT improved with keywords from a controlled list or from a specific thesaurus or with a classification system. Explain how you can apply in this case a thesaurus to improve your searches.
Knowledge organisation: relations among some tools

- Ontologies / Topic maps
- Thesauri
- Controlled vocabularies
Knowledge organisation: classifications and thesaurus systems

Classification systems
versus
thesaurus systems
Knowledge organization: classifications versus thesauri

• Classification
  » Good for placement of documents in a library (because documents on many related subjects can be kept together)
  » Not well suited for computer searching (too complicated)

• Thesaurus
  » Not suited for placement of documents in a library (because documents with related subjects would NOT be kept together)
  » Well suited for computer searching (relatively simple alphabetic listing of keywords)
• You are free to copy, distribute, display this work under the following conditions:

» Attribution:
   You must mention the author.

» Noncommercial:
   You may not use this work for commercial purposes.

» No Derivative Works:
   You may not change, modify, alter, transform, or build upon this work.

• For any reuse or distribution, you must make clear to others the license terms of this work.