Knowledge organisation: classifications and thesaurus systems

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
Give examples of general, universal classification systems.

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.
» ...

?? Question ??

When people search to find, can classification help to increase recall? If yes, how?
Consider for instance
- A supermarket
- A classical book library
- The WWW
When people search to find, can classification help to increase *precision*? If yes, how?

Consider for instance:
- A classical book library
- The WWW

The taxonomy/classification of biological species is a good and well-known example of a taxonomy/classification.

However, it is nevertheless an exceptional and not a typical classification system like the ones used to classify information items.

*Explain this paradox.*
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
### Thesaurus relations

<table>
<thead>
<tr>
<th>Term(s) with broader meaning</th>
<th>Term(s) with narrower meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Term</strong></td>
<td><strong>Term</strong></td>
</tr>
<tr>
<td><strong>Other term(s)</strong></td>
<td><strong>Synonym(s)</strong></td>
</tr>
</tbody>
</table>

?? Question ??

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)

Term(s) with broader meaning

- BT (= Broader Term)
- RT (= Related Term)
- Other term(s)
- Term
- Synonym(s)

Term(s) with narrower meaning

- NT (= Narrower Term)
- UF (= Use(d) For)
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).
Thesaurus applications

- To find more and/or better terms during writing.
- To understand the meaning of a term, by inspecting
  - the scope note of the term and/or
  - the relations with other terms.

!! Task - Assignment !!

Read the book chapter about
Language and information retrieval
by Large, Andrew, Tedd, Lucy A., and Hartley, R.J.
In: Information seeking in the online age:
principles and practice.
!! Task - Assignment !!

Read the book chapter about
Language in information representation and retrieval.
by Chu, Heiting
in Information representation and retrieval in the digital age.
ASIST Monograph Series.

**--

?? Question ??

Which thesauri do you know?
Thesaurus systems that cover all subjects

- General systems
- Universal systems
- Covering all subjects
- Broad and shallow systems
- Horizontal 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/
Thesaurus systems on the WWW that cover all subjects: 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 through the WWW: screenshot
General thesaurus system through the WWW: screenshot *sea*

General thesaurus system through the WWW: screenshot *ocean*
Thesaurus systems covering all subjects: comments

- An ideal, complete 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.
!! Task - Assignment - Exercise !!

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.

Have a look at various global, general, universal thesaurus systems. Consider which ones may be useful for your future online information searches.
Thesaurus systems focused on a particular subject

- Focused on a particular subject domain = narrow and deep, vertical systems

Examples

- **ERIC**: education, information science, ...
- *Psychological Abstracts / PsycInfo*
- *Sociological Abstracts / SocioFile*
- **INSPEC**: physics, electronics, information technology
- the *Aquatic Sciences and Fisheries Information System*
- *Medline* (the Medical Subject Headings = MeSH)
- Various thesaurus systems for art and architecture can be found online: [http://www.getty.edu/research/tools/vocabulary/](http://www.getty.edu/research/tools/vocabulary/)
**--Examples

Thesaurus systems focused on a particular subject: examples

- A database of thesaurus systems is accessible online through [http://www.taxonomywarehouse.com/](http://www.taxonomywarehouse.com/)

---

***--

?? Question ??

Give an example of a horizontal thesaurus for the whole English language and of a vertical thesaurus for a particular subject domain.
**Question**

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

- Controlled vocabularies
- Thesauri
- Ontologies / Topic maps
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)
Use the *Aquatic Sciences and Fisheries Thesaurus* through the Internet ([http://www4.fao.org/asfa/asfa.htm](http://www4.fao.org/asfa/asfa.htm)) to find the appropriate terms to retrieve items about “fishing with poison” from the database of the *Aquatic Science and Fisheries Information System*.

**Example 41**

Use the *Aquatic Sciences and Fisheries (ASFA) Thesaurus* to formulate a query to find general reviews about monitoring of sea pollution, in the database of the *Aquatic Science and Fisheries Information System*.

**Example 42**
• 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.