

Working group 4

Problems and prospects in thesaurus construction

Reported, arranged, and edited by

Dagobert Soergel

The discussion identified the following issues

1. Thesaurus use in relation to structure

- The age-old question: Do we need thesauri and do they really improve retrieval? Or more precisely: Under what circumstances can a good thesaurus make what difference in retrieval and use of information? Which elements of thesaurus structure have what effects? Still more research is needed, taking into consideration improved methods in the structuring and application of thesauri
- How do users interact with thesauri, seen as a function of user knowledge (for example, user knowledge of using hierarchy in searching and of specific hierarchies). This needs more research, taking into consideration the possibilities of users interacting with electronic thesauri using improved interfaces and search possibilities
- Interfaces for interacting with thesauri, especially for end users on the Web
Integrated interface to multiple thesauri and other lexical resources
- Behind the scenes use of thesauri to improve retrieval.
A thesaurus as a knowledge base for a user-system dialog
Use of thesaurus structure to support browsing
- User learning from thesaurus structure.
Thesaurus as a resource for users in interpreting information, for example by providing concepts and terminology of a domain unfamiliar to the user
Thesaurus as an education tool generally
- Effect of thesaurus structure and content on user success. Which thesaurus elements influences success
- Application of thesauri to indexing very large collections, specifically the Web.
Thesauri as knowledge bases for automated indexing
- Use of thesauri in collection building (if the object in question cannot be indexed with the thesaurus, either the object is out of scope or the thesaurus needs to be updated)
- Thesaurus structure in relation to the granularity of the objects retrieved (whole documents, small units of text, whole data tables, individual data such as property data or commodity prices)

2. Thesaurus structure and content

- Is hierarchy needed?
- Types of relations. Need richer set of relationships
- Multilingual thesauri. Linking terms across languages
Relationship to linguistic approaches to information retrieval
- Cross-disciplinary or multi-domain thesauri. Linking terms across domains
Tension between usefulness of a domain framework and desire to search multiple domains
- Developing a shared vocabulary for access to images
- Visual elements (signs and icons, shapes, pictures, structure diagrams) as thesaurus elements

3. Evaluation of thesauri

4. Thesaurus construction

- Knowledge-based methods (use of linguistic/semantic knowledge and analysis) vs statistical methods in thesaurus construction. Combining methods. Manual vs automated implementation of methods or a combination.
- Support for the construction of large scale thesauri - automated
- Correlating existing thesauri for both search support and support in thesaurus construction
Correlating thesauri from different disciplines/domain (see also 2 above)
Correlating thesauri with different structure. How to compare hierarchies
- Adaptation of existing classifications and terminologies to an electronic environment (see also interface under 1)
- Thesaurus construction in the absence of a collection. How to anticipate need for concepts and terms in the thesaurus
- Dynamic thesaurus construction and update
User input
Query analysis
New concepts and terms in documents, detected automatically or communicated by indexers