Dagobert Soergel
College of Library and Information Services
University of Maryland, College Park
ds52@umail.umd.edu

Thesauri for knowledge-based assistance in searching digital libraries

DL97 Workshop on Thesauri and Metadata

Challenges for hypermedia and digital libraries

Improve retrieval effectiveness to handle the sheer mass of material

Provide unified access to materials in different media (esp. access to non-text materials)

Provide knowledge-based support for end users who access networked information without the benefit of an intermediary

Support creation and maintenance of personal or work-group information systems

Support information seeking as an integral part of problem solving, learning, and intellectual work

Support collaborative work:
Scholarly communication as computer-supported multi-party conversation

Metadata must support these functions

Support information seeking as an integral part of problem solving, learning, and intellectual work

Help users to explore ideas in conjunction with exploring information

Support fine-grained retrieval and assimilation of information

Support processing of information along with or after retrieval

Support collaborative work

Make users full participants in the continuing improvement of information systems through feedback and other contributions

Establish linkages between people

The structure of metadata

Types of data included - defined in an entity-relationship model expressed through fields, tags, or property-value pairs

Entity values or field content.
Structure of scheme of values, as in a thesaurus or classification

Functions of a thesaurus / classification / ontlogical knowledge base

in the context of digital libraries

Support learning and assimilating information.

Assist researchers and practitioners with problem clarification.

Support information retrieval.

- **Provide knowledge-based support for end-user searching.**
- **Support meaningful information display.**
- **Provide a tool for indexing.**
- **Facilitate the combination of multiple databases or unified access to multiple databases.**

Support document processing after retrieval.

Support learning and assimilating information

Support learning about any topic by providing the learner with a coherent, age-appropriate conceptual framework.

Learning as information retrieval. Conceptual framework for asking the right questions.

Assist readers in understanding text.

Assist researchers and practitioners with problem clarification —

provide the conceptual basis for the design of good research and implementation and for good query formulation. Includes help with

exploring the conceptual context of a research or practical problem — a study, policy, plan, or implementation project

and with

structuring the problem.

Assist researchers and practitioners with problem clarification

Examples of specific functions:

Present the issues in a field or application area in a coherent framework.

Assist in problem-solving: Assist in the exploration of the dimensions of a problem and aspects to be considered in its solution; provide a classification of approaches to solving a specific problem.

Provide classification and consistent definition of variables for research / of evaluation criteria for practical problems, thus enhancing the comparability of research and evaluation results and making research more cumulative.

Support information retrieval

Provide knowledge-based support for end-user searching. Support

- **searching in multiple natural languages;**
- **free-text searching;**
- **searching multiple databases using different index languages.**

Knowledge-based search support, continued

Elicitation of user needs through a series of menus based on search tree, or through guidance in the conceptual analysis of a search topic (questions based on a facet structure, presentation of a segment of the concept hierarchy for each applicable facet).

Browsing the classification structure to identify useful concepts for a search at the level of specificity desired.

Mapping from the user's query terms to descriptors used in a database or to the multiple natural language expressions to be used for free-text searching.

Inclusive (hierarchically expanded) searching.

Enhanced ranking algorithms based on concept and term relationships.

Searching multiple databases by mapping the users query terms to the descriptors used in each of the databases, or mapping the descriptors from one database to another databases (switching); common search language.

Support information retrieval, continued

Support information display, especially presentation of search results:

Meaningful arrangement of units (document records, paragraphs, property data on a given substance assembled from several databases), including knowledge-based clustering of records retrieved.

This supports exploration of large retrieved sets and, by extension, exploration of the content of an entire collection or subcollection.

Meaningful arrangement of information within a record (for example meaningful ordering of descriptors assigned).

Support information retrieval, continued

Provide a tool for indexing.

Vocabulary control.

User-centered (request-oriented, problem-oriented) indexing.

Indexing several databases in a field with a common index language and sharing the results of indexing to reduce overall indexing effort.

Mapping indexing descriptors from one system to another.

Support information retrieval, continued

Facilitate the combination of multiple databases or unified access to multiple databases through

- **mapping the users query terms to the descriptors used in each of the databases;**
- **mapping the query descriptors from one database to another (switching);**
- **providing a common search language from which to map to multiple databases;**
- **providing a common index language for a number of databases in a field;**
- **mapping indexing descriptors from one database to another.**

Support document processing after retrieval

For example

Highlight descriptors responsible for retrieval, using different colors for different facets.

Highlight terms belonging to a given category, for example, personal names, again using different colors for different categories.

Prepare document summaries, possibly in a different language , taking into account the query topic.

Translate full documents.

Extract facts from text. Compile and arrange facts extracted from several texts.

The underlying function of a knowledge base on concepts and terminology

Map out a concept space, relate concepts to terms, and provide definitions, thus providing orientation and serving as a reference tool.

Provide a semantic road map and common language for an individual field and, perhaps more importantly, map the relationships among fields.

Clarify concepts by putting them in the context of a classification / typology and to provide a system of definitions.

Relate concepts and terms across disciplines, languages, and cultures.