Functions of a thesaurus / classification / ontological knowledge base

College of Library and Information Services
University of Maryland
October 1997
## Functions of a thesaurus / classification / ontological knowledge base

### Overview

Provide a **semantic road map** to individual fields and the relationships among fields. Map out a concept space, relate concepts to terms, and provide definitions, thus providing orientation and serving as a reference tool.

**Improve communication generally. Support learning and assimilating information.**

- Support learning through conceptual frameworks. Conceptual framework to help the learner ask the right questions.
- Support the development of instructional materials through conceptual frameworks.
- Assist readers in understanding text by giving the meaning of terms.
- Assist writers in producing understandable text by suggesting good terms.
- Support foreign language learning.

**Provide the conceptual basis for the design of good research and implementation.**

Assist researchers and practitioners with problem clarification.

**Provide classification for action**

- a classification of diseases for diagnosis,
- of medical procedures for insurance billing,
- of commodities for customs.

**Support information retrieval.**

- Provide a tool for searching, particularly knowledge-based support for end-user searching, including hierarchically expanded searching.
- Provide a tool for indexing.
- Facilitate the combination of multiple databases or unified access to multiple databases.
- Support document processing after retrieval.

**Support meaningful, well-structured display of information.**

**Ontology for data element definition.** Data element dictionary.

**Conceptual basis for knowledge-based systems.**

**Do all this across multiple languages**

**Mono-, bi-, or multilingual dictionary for human use.**

**Dictionary/knowledge base for automated language processing**
The underlying function of a knowledge base on concepts and terminology:

Provide a semantic road map to individual fields and the relationships among and across fields.

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.

Many specific functions build on this foundation.
Improve communication generally.
Support learning and assimilating information

**Support learning** about any topic by providing the learner/reader with a **coherent, age-appropriate conceptual framework**. Conceptual frameworks help the learner ask the right questions; **learning as information retrieval**.

**Support the development of instructional materials** by providing a conceptual framework to the instructional developer / writer and by suggesting didactically useful arrangements of topics.

**Assist readers** in understanding text; help them ascertain the proper meaning of a term and placing it in context.

**Assist writers** in producing understandable text by helping them to conceptualize the topic and suggesting from a semantic field the term that best conveys the intended meaning and connotation.

**Support foreign language learning**
Provide the conceptual basis for the design of good research and implementation.

Assist researchers and practitioners with problem clarification

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 and providing a conceptual framework for asking the right questions and devising good query formulations for retrieval.

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 (for example, a classification of approaches to drug abuse prevention as a help in designing drug abuse prevention projects).

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.
## Provide classification for action

For example,

- a classification of diseases for diagnosis,
- a classification of medical procedures for insurance billing,
- a classification of medical outcomes to assist with treatment evaluation,
- a classification of commodities for customs,
- a classification of educational objectives for instructional development,
- a classification of occupations for matching job applicants with job openings and for pay scale;
- a classification of skills for employee task assignments.
Support information retrieval 1:

A tool for searching, particularly knowledge-based support for end-user searching. Support searching in any kind of database — bibliographic, full-text and hypermedia, directory, numeric, etc.;

searching in any kind of medium — printed indexes, CD-ROM systems, online systems, and the Internet;

searching in multiple natural languages independent of the language used in each database;

free-text searching;

searching multiple databases using different index languages.

Elicitation of user needs through a series of menus based on a 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. (The user may not have command of the vocabulary needed.)

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 that use 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 2:

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 3:

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 information retrieval 4:

Support document processing after retrieval

Sample functions that require knowledge-based support:

- **Meaningful arrangement of search results** (see next box)

- **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 substantive data from text**. Compile and arrange data extracted from several texts.

Support meaningful, well-structured display of information

- **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 unit** (for example meaningful ordering of descriptors within a bibliographic record).
Ontology for data element definition.

Data element dictionary.

Consider data processing systems in a multinational corporation

Conceptual basis for knowledge-based systems.

Do all this across multiple languages

Mono-, bi-, or multilingual dictionary for human use.

Printed or machine-readable, such as dictionary on CD-ROM or a thesaurus used in conjunction with a word processor

Dictionary/knowledge base for automated language processing

Machine translation and natural language understanding (data extraction, automatic abstracting/indexing). (It should be noted that parsing natural language requires not only morphological information and information about the possible syntactic roles of a term but also a great deal of semantic information.)

Spell check dictionary

Knowledge base for grammar checking.
Functions of an ontological knowledge base in software development

Assist in the design and implementation of the **user interface, esp. choice of terms and icons**.

Terms and icons must be chosen with the sometimes conflicting goals of communicating to the intended user group and of adhering to standards.

Assist in the organization and formulation of **help messages and of documentation** and third-party software books.

Serve as the **lexicon for machine translation** of interfaces and software-related documents.

**Assist the user in understanding interfaces and documentation, esp. in a foreign language.**

**Support retrieval of software** for the end user or for **software reuse**.

**Data element definition and standardization and organization of CASE tool databases.**

All this functionality must be provided in **multiple languages** (for example, **software localization** for end users, **CASE tool databases for multinational development teams**).
**Illustrative thesauri / classification schemes**

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bloom</td>
<td>Taxonomy of educational objectives</td>
<td>(1 copy in the cataloging laboratory) (LibSch LB17.B55.1956)</td>
</tr>
<tr>
<td>WICHE</td>
<td>Management Information Systems Program; Data element dictionary</td>
<td>(1 copy - 5 v. in cataloging lab, very thin light green books) (LibSch Z699.5.E3M3)</td>
</tr>
<tr>
<td>ERIC</td>
<td>Education Resources Information Center Thesaurus. 13th ed.</td>
<td>3 copies in the Cat. Lab. (brown) In a pinch, the 12th edition, 1990 (purple), will do (also in Cat. Lab)</td>
</tr>
<tr>
<td>MeSH</td>
<td>Medical Subject Headings</td>
<td>(The published edition 1990 or later is fine for purposes of this assignment.) LibSch Ref Z695.1. M48U5 (1996).</td>
</tr>
<tr>
<td>AOD</td>
<td>The Alcohol and Other Drug Thesaurus.</td>
<td>2nd ed., 1995.</td>
</tr>
<tr>
<td>Yahoo</td>
<td>The Yahoo classification</td>
<td></td>
</tr>
<tr>
<td>WordNet</td>
<td>WordNet, a vocabulary and classification database accessible through the Web</td>
<td>(<a href="http://www.notredame.ac.jp/cgi-bin/wn.cgi">www.notredame.ac.jp/cgi-bin/wn.cgi</a>)</td>
</tr>
<tr>
<td>HS</td>
<td>Harmonized Commodity Description and Coding System.</td>
<td>World Customs Organization, Brussels. info: <a href="http://pacific.commerce.ubc.ca/trade/HS.html">http://pacific.commerce.ubc.ca/trade/HS.html</a></td>
</tr>
</tbody>
</table>
NAICS

**North American Industrial Classification System**

"provides common industry definitions for Canada, Mexico, and the United States. It was developed in cooperation with the US Economic Classification Policy Committee, Statistics Canada, and Mexico's Instituto Nacional de Estadistica, Geografia e Informatica to better compare economic and financial statistics and ensure that such statistics keep pace with the changing economy. The new NAICS system will replace the countries' separate classification systems with one uniform system for classifying industries. In the United States, NAICS will replace the Standard Industrial Classification system."


ICD-10

**The International Statistical Classification of Diseases and Related Health Problems, tenth revision.** Produced by the World Health Organization. Published in many languages.

For info: www.who.org/programmes/hst/icd-10/icd-10.htm


CPT


Health Care Finance Administration (HCFA) Common Procedure Coding System (HCPCS) is the basis for Medicare reimbursement for hospital outpatient services. It is comprised of three levels - CPT (level 1), HCPCS or National (level 2), and Local (level 3).

In its data collection the Agency for Health Care Policy and Research (AHCPR) uses data standards that are based on those employed by the Census Bureau, the American Hospital Association, the Health Resources and Services Administration (Area Resource File), the National Center for Health Statistics, and codes for clinical diagnosis and procedures such as ICD-9, ICD-9-CM, and CPT-4. These standards are intended to facilitate data analysis and use by ensuring comparability, quality and interoperability. Further, by promoting uniform, accurate, and automated health care data, AHCPR advances medical research (including medical effectiveness and cost effectiveness research) and improves the efficiency of the private sector health care delivery system and quality improvement measurement.

A further type of classification are **biological taxonomies.** There are several rivaling schemes for major areas (kingdoms) and many publications on specific areas. They are used in biology but also in agriculture, food science, and medicine.