

TermMaster. Software for creating and maintaining thesauri and ontologies

The construction and maintenance of thesauri and ontologies is complex business. It requires a database of many sources to exploit existing intellectual capital, the maintenance of a complex network of relationships of many types, distinction of application-specific subsets within an organization-wide thesaurus or ontology, and excellent facilities to produce meaningful presentations in print and on the Web. TermMaster, a legacy program developed over the years for the Alcohol and other Drug Thesaurus (<http://etoh.niaaa.nih.gov/AODVol1/Aodthome.htm>) supports these functions with a rich set of features; write to ds52@umail.umd.edu for a free pre-beta version (for non-commercial purposes). But be forewarned: the program was developed for use by a small number of knowledgeable staff; it rates high on power but low on usability; it runs under DOS, and all functions are invoked through configuration files. For a description and introductory user manual see <http://www.clis.umd.edu/faculty/soergel/dlthestut/SoergelTermMasterDoc.PDF>.

The following is a description of the more important capabilities of TermMaster.

General

Stores **multiple thesauri/ontologies in an integrated thesaurus data base** and tracks the origin(s) of each piece of information. Supports the construction of a new thesaurus/ontology based on several source thesauri. To facilitate comparisons, TermMaster stores terms as stems, with the appropriate singular or plural suffix given with each source in which the term occurs; where singular and plural have different meanings, this can be overridden.

Supports the **creation, maintenance, and meaningful display of hierarchies**.

Allows a **large and extensible number of relationship types**, including different types of scope notes.

Data Input

Accepts input files in several formats (produced with a word processor or by conversion from a machine-readable source), mainly hierarchical and alphabetical. Preserves the hierarchical structure and sequence of input files..

Does extensive input checking to assure formal correctness of input files.

Data Manipulation and Editing

Creates external term numbers (notations) expressing the hierarchy, building on “seed numbers” given for terms on the upper levels of the hierarchy.

Creates hierarchical cross-references implied by a hierarchic input file.

Checks thesaurus terms embedded in scope notes (marked by * __ *), replaces any nonpreferred term with the descriptor, and adds the term number.

Comparison of Thesauri

Creates thesaurus comparison files that support the mapping from one thesaurus to another and the creation of complete thesauri. One option for a comparison file lists candidate terms that occur in one or more source thesauri but not in the target thesaurus under construction. Another option is to import conceptual relationships from any number of sources into a target thesaurus even if different thesauri use different terms to express these relationships. (TermMaster uses synonym relationships found in the target thesaurus and in the sources to accomplish this.)

Output Capabilities: Print, Web, export files

Print formats

Well-designed hierarchical displays (outline or entire hierarchy), giving the user complete control over formatting and over the information to be printed for each term: the user can specify groups of internal relationships to be treated as one external relationship (e.g., map AB, FT, ST, and ET to ST) and specify the sequence of these groups. One can print a quick hierarchy that lists just descriptors or an annotated hierarchy that gives scope notes and cross-references.

Alphabetical index of all terms (descriptors and nondescriptors) in KWOC format.

Annotated alphabetical list of all terms, if desired.

Web formats

HTML pages that maintain the user-friendly format of the printed pages, including the indented hierarchical arrangement. Cross-references are all active hyperlinks.

Input file for DB/TextWorks to set up a searchable thesaurus database on the Web. The descriptor records found are hyperlinked into the appropriate HTML thesaurus page so that the user can see a retrieved descriptor in its hierarchical context.

Data Export Files

Files in the native format of UMLS (Unified Medical Language System) and a delimited ASCII file of descriptors and their term numbers for input into a database management system.