

UMLS[®] Reference Manual

U.S. National Library of Medicine

Introduction to the UMLS

[7]

1.1 Purpose of the UMLS

The Unified Medical Language System (UMLS) facilitates the development of computer systems that behave as if they "understand" the language of biomedicine and health. To that end, NLM produces and distributes the UMLS Knowledge Sources (databases) and associated software tools (programs). Developers use the Knowledge Sources and tools to build or enhance systems that create, process, retrieve, and integrate biomedical and health data and information. The Knowledge Sources are multi-purpose and are used in systems that perform diverse functions involving information types such as patient records, scientific literature, guidelines, and public health data. The associated software tools assist developers in customizing or using the UMLS Knowledge Sources for particular purposes. The Lexical Tools work more effectively in combination with the UMLS Knowledge Sources, but can also be used independently.

1.2 Conditions of Use of the UMLS

All UMLS Knowledge Sources and associated software tools are free of charge to U.S. and international users.

The Semantic Network, the SPECIALIST Lexicon, and associated Lexical Tools are accessible on the Internet under open terms, which include appropriate acknowledgment for their use. View the terms and conditions for use of the Semantic Network and of the SPECIALIST Lexicon and Lexical Tools.

To use the Metathesaurus, you must establish a license agreement. This is because the Metathesaurus includes vocabulary content produced by many different copyright holders as well as the substantial content produced by NLM.

Setting up the license agreement is done via the Web. Once the license agreement is in place, much of the content of the Metathesaurus may be used under very open conditions. Your pre-existing licenses for content with use restrictions, e.g., CPT, MedDRA, or NIC, will cover your use of that content as distributed within the Metathesaurus. Some vocabulary producers who require authorization to use their content will generally grant free permission.

The complete text of the License Agreement for Use of the UMLS Metathesaurus appears in Appendix A of the current UMLS release documentation.

1.3 The UMLS Knowledge Sources and Associated Tools

There are three UMLS Knowledge Sources: the Metathesaurus, the Semantic Network, and the SPECIALIST Lexicon. They are distributed with several tools that facilitate their use, including the MetamorphoSys install and customization program.

1.3.1 Metathesaurus

The Metathesaurus is a large, multi-purpose, and multi-lingual vocabulary database that contains information about biomedical and health-related concepts, their various names, and the relationships among them. It is built from the electronic versions of numerous thesauri, classifications, code sets, and lists of controlled terms used in patient care, health services

billing, public health statistics, indexing biomedical literature, and/or basic, clinical, and health services research. In this documentation, these are referred to as the "source vocabularies" of the Metathesaurus. In the Metathesaurus, all the source vocabularies are available in a common, fully-specified database format.

A complete list of the source vocabularies present in the current version of the Metathesaurus appears in Source Vocabularies page of the current UMLS release documentation. The list indicates which coding systems and vocabularies are designated as U.S. standards for administrative health transactions in accordance with HIPAA or as target U.S. government-wide clinical standards selected by the Consolidated Health Informatics eGov initiative.

The Metathesaurus is organized by concept or meaning. In essence, it links alternative names and views of the same concept and identifies useful relationships between different concepts. All concepts in the Metathesaurus are assigned at least one Semantic Type from the Semantic Network (1.3.2) to provide consistent categorization at the relatively general level represented in the Semantic Network. Many of the words and multi-word terms that appear in concept names or strings in the Metathesaurus also appear in the SPECIALIST Lexicon (1.3.3). The Lexical Tools are used to generate the word, normalized word, and normalized string indexes to the Metathesaurus. MetamorphoSys (1.3.5) is used to install the UMLS Knowledge Sources and customize the Metathesaurus.

The Metathesaurus *must* be customized to be used effectively.

A complete description of the Metathesaurus and its file structure begins with Chapter 2 of this documentation.

1.3.2 Semantic Network

The Semantic Network provides a consistent categorization of all concepts represented in the Metathesaurus and provides a set of useful relationships between these concepts. All information about specific concepts is found in the Metathesaurus; the Network provides information about the set of basic Semantic Types, or categories, which may be assigned to these concepts, and it defines the set of relationships that may hold between the Semantic Types. The Semantic Network contains 135 Semantic Types and 54 relationships. The Semantic Network serves as an authority for the Semantic Types that are assigned to concepts in the Metathesaurus. The Network defines these types, both with textual descriptions and by means of the information inherent in its hierarchies.

The Semantic Types are the nodes in the Network, and the Semantic Relations between them are the links. There are major groupings of Semantic Types for organisms, anatomical structures, biologic function, chemicals, events, physical objects, and concepts or ideas. The current scope of the UMLS Semantic Types is quite broad, allowing for the semantic categorization of a wide range of terminology in multiple domains.

A complete description of the Semantic Network and its file structure appears in Chapter 5 of this documentation.

1.3.3 SPECIALIST Lexicon and Lexical Tools

The SPECIALIST Lexicon is intended to be a general English lexicon that includes many biomedical terms. Coverage includes both commonly occurring English words and biomedical vocabulary. The lexicon entry for each word or term records the syntactic, morphological, and orthographic information needed by the SPECIALIST Natural Language Processing System.

The Lexical Tools are designed to address the high degree of variability in natural language words and terms. Words often have several inflected forms which would properly be considered instances of the same word. The verb "treat", for example, has three inflectional variants:

- treats — the third person singular present tense form
- treated — the past and past participle form
- treating — the present participle form

Multi-word terms in the Metathesaurus and other controlled vocabularies may have word order variants in addition to their inflectional and alphabetic case variants. The Lexical Tools allow the user to abstract away from several types of variation, including British English/American English spelling variation and character set variations.

A complete description of the SPECIALIST Lexicon, its file structure, and the lexical programs appears in Chapter 6 of this documentation.

1.3.4 UMLS Knowledge Source Server

The UMLS Knowledge Source Server (UMLSKS) is a set of Web-based interactive tools and a programmer interface that allows users and developers to access the UMLS Knowledge Sources, including the vocabularies within the Metathesaurus. It also contains the download site for the UMLS data files. The UMLSKS is a useful starting point for gaining an understanding of the content of the UMLS resources. Because it contains the complete Metathesaurus files, access to UMLSKS is restricted to registered users who have signed the License Agreement for Use of the UMLS Metathesaurus.

A complete description of the UMLS Knowledge Source Server and its capabilities appears in Chapter 7 of this documentation.

1.3.5 MetamorphoSys: The UMLS Installation and Customization Program

MetamorphoSys is a cross-platform Java application that must be used if the UMLS Knowledge Sources (Metathesaurus, Semantic Network, and SPECIALIST Lexicon) are installed locally. MetamorphoSys also supports the creation and refinement of customized subsets of the Metathesaurus. In general, the Metathesaurus must be customized to be used effectively in specific applications.

MetamorphoSys guides you first through the installation of one or more UMLS Knowledge Sources, and then through customization of the Metathesaurus for local use. A variety of options are available, such as the inclusion or exclusion of specific source vocabularies, languages, and term types, specification of output character set (7-bit ASCII or Unicode UTF-8) and output format (Rich Release Format or Original Release Format) for the Metathesaurus files.

A complete description of MetamorphoSys appears in Chapter 8 of this documentation.

1.4 Getting Started

The UMLS resources are powerful - and unusual - tools intended for use by system developers. Here are a few suggestions about how to start building your understanding of UMLS features and capabilities and their potential for enhancing your applications.

Scan the entire UMLS documentation to get a sense of the range of resources available.

If the Metathesaurus interests you, take time to read Chapter 2 of the documentation. The background there will make it easier to understand the actual file descriptions in Chapter 3 and Chapter 4.

Use the Web registration system to execute the free License Agreement for Use of the UMLS Metathesaurus. A license agreement is required because the Metathesaurus contains vocabularies produced by many different copyright holders. You are able to use much of the content of the Metathesaurus with minimal restriction, but you may need to obtain additional licenses from individual vocabulary producers if you wish to use certain vocabularies contained in the Metathesaurus. The various restriction levels are explained in the UMLS license agreement and its Appendix.

Once you have executed the current license agreement, use the UMLS Knowledge Source Server for initial browsing and exploration of the contents of the Metathesaurus, Semantic Network, and SPECIALIST Lexicon and of additional special resources useful to application developers.

If you require local copies of the UMLS files, use the MetamorphoSys install and customization program described in Chapter 8 of this documentation to produce them. You may find it useful to experiment with various options to produce customized subsets. MetamorphoSys comes on the UMLS DVD and is available for download with the UMLS data files from the UMLS Knowledge Source Server.

1.5 Sources of Additional Information about the UMLS

In addition to providing links to the UMLS documentation and to the UMLS Knowledge Source Server, the UMLS Web site links to fact sheets on the UMLS Knowledge Sources and Knowledge Source Server; FAQs; training materials; and information about NLM applications and research projects that use the UMLS. Articles on the UMLS project and resources can be retrieved from MEDLINE/PubMed. Click [here](#) to obtain a current search. A comprehensive 1986-1996 bibliography on the UMLS project covering additional papers not indexed for MEDLINE/PubMed is also available.

UMLS users are strongly encouraged to subscribe to the UMLS users listserv. NLM uses the listserv to seek advice from users and to distribute news about upcoming UMLS developments; users share experiences or obtain advice about using the UMLS resources.

To subscribe, send an e-mail to listserv@list.nih.gov containing the following message:
SUBSCRIBE UMLSUSERS-L <your full name>.

To unsubscribe, send an e-mail to listserv@list.nih.gov containing the following message:
SIGNOFF UMLSUSERS-L <your full name>.

To post a message to the list AFTER subscribing, send an e-mail to UMLSUSERS-L@list.nih.gov.

To access subscription information and list archives, go to [UMLSUSERS-L Listserv Web page](#).

Specific questions about the UMLS can be addressed to NLM Customer Service or, for telephone inquiries, to 1-888-FINDNLM (1-888-346-3656).