01/01/2002

UMLS KNOWLEDGE SOURCES
13th Edition - January Release
2002AA
DOCUMENTATION

SECTION 0.
INTRODUCTION TO THE UMLS

0.1 What's New for the 2002AA UMLS

* Notice: 2002 APPENDIX TO THE LICENSE FOR USE OF THE UMLS PRODUCTS

Included in the package is a copy of the 2002 Appendix to the License Agreement for the UMLS Products. The License Agreement and the current Appendix are also available on NLM's Web site at:

http://www.nlm.nih.gov/research/umls/license.html

Please read the Appendix carefully. Continued use of the UMLS products under the terms of your original License requires that you have read the information in the 2002 Appendix to the License for Use of the UMLS Products.

Continued use of the UMLS means that you agree and accept that the annually updated Appendix is part of the License Agreement and that you are responsible for adhering to any additional restrictions on use of some of the source vocabularies contained in the Metathesaurus that are described in the Appendix. An explanation of the meaning of the various categories of restrictions appears in the body of the license agreement.

0.1.0 Introduction

Please note that the UMLS Knowledge Sources are created for developers and are not end-user applications. Significant intellectual effort and computing resources may be required to understand and use them correctly!

E-mail questions about the UMLS should go to umls_support@nlm.nih.gov. Telephone questions go to 1-888-FINDNLM.

Here are four further resources for UMLS information:

1.) UMLS Knowledge Source Server: http://umlsks.nlm.nih.gov

The UMLS Knowledge Source Server provides the full UMLS data. It is open to anyone who has signed the UMLS License Agreement (http://www.nlm.nih.gov/research/umls/license.html) and has received a license number. First time users should establish a login and a password through the online registration at the web site. Any questions or problems should be addressed via email to umlsks@nlm.nih.gov.

2.) Umlsinfo: http://umlsinfo.nlm.nih.gov

This site provides helpful information - documentation, FAQs, learning resources including tutorials, bonus files, sample load scripts, and more. It is an evolving resource; please give us feedback. This site does not contain proprietary information and so does not require a UMLS license.

3.) Umls-users listserv

UMLS users can subscribe to a listserver to discuss UMLS issues. To subscribe to the listserver, simply send a message to listserver@nlm.nih.gov which includes the following line:

subscribe umls-users

To post a message to the umls-users listserver AFTER subscribing, send email to umls-users@lhc.nlm.nih.gov

4.) Lexical Tools documentation is available at:
http://umlslex.nlm.nih.gov/lvg/docs/userDoc/index.html

Release Names:

To allow for multiple releases each year, releases are named in the format: YYYYVV_SS, where:

- YYYY is a 4 digit year, e.g., 2002

- VV is a two letter designation for the version ID starting with "AA", e.g., 2002AA

- SS is a two digit subset ID where needed (e.g., "02")

The first release of the year (called a January release) will have the version ID of "AA". Subsequent releases for that year will have a version ID of "AB", "AC", etc.

Releases that are created as subsets using MetamorphoSys should have a subset ID designation starting with 01 and counting up.

Examples:

- 2002AA -> January release for 2002

- 2002AC -> third release (AC) for 2002

Compact Disc Names:

Discs are named for release and type, for example:

2002AA_ZIP

2002AB_TAR

2002AC_02_TAR

(Note: these names will work even when they include the subset number, e.g. 2002AC_02_TAR)

Compressed file names:

The ZIP and TGZ compressed files are named for the release with a ZIP or TGZ extension, for example:

2002AB.ZIP

2002AC_02.TGZ

0.1.1 Metathesaurus

For 2002: MORE names, Fewer Concepts!

The 2002AA edition of the Metathesaurus includes 776,940 concepts and 2.10 million concept names in its source vocabularies. Compared with the 2001AA version, there were 20,419 fewer concepts, yet there are 137,056 more strings in 2002AA MRCON and 185,464 more names in MRSO. There are approximately 11,137,725 relationships, an increase of 458,219.

New for the 2002AA Metathesaurus are: DUT2001, Dutch Translation of Medical Subject Headings; MDR40, Dictionary for Regulatory Activities Terminology (MedDRA) version 4.0; MTHMST2001 - Minimal Standard Terminology Digestive Endoscopy (English, French and Italian versions); VANDF01, Veterans Health Administration National Drug File.

Twenty-six sources were updated: BRMP2002 - Portuguese translation of MeSH; BRMS2002 - Spanish translation of MeSH; CDT3 - Current Dental Terminology; CPT2002 - Physicians' Current Procedural Terminology; CPT01SP - Spanish version of CPT; CSP2001 - CRISP; DMD2002 - German translation of MeSH; FIN2002 - Finnish translation of MeSH; HCDT3 - HCPCS Version of Current Dental Terminology; HCPCS02 - Health Care Financing Administration Common Procedure Coding System; HCPT02 - HCPCS Version of Current Procedural Terminology; ICD2002 - ICD-9-CM. 6th ed.; ITA2002 - Italian translation of MeSH; LNC203 - Logical Observation Identifier Names and Codes; MMSL01 - Multum MediSource Lexicon; MMX01 - Micromedex; MSH2002 - Medical Subject Headings (MeSH); MTHCH02 - UMLS Metathesaurus Hierarchical CPT terms; MTHHH02 - UMLS Metathesaurus Hierarchical HCPCS terms; NCI2001a - NCI Thesaurus; NDDF01 - First Databank's National Drug Data File;

PDQ2001 - PDQ; PSY2001 - PsycInfo; RUS2002 - Russian translation of MeSH; UMD2002 - UMDNS; UWDA155 - Digital Anatomist version 1.5.5.

The first-ever DECREASE in the number of Concepts is primarily due to Concept level review of the previously "Unreviewed" MeSH Supplementary Concepts in the new MeSH system. There have been other efforts to find missed synonymy as well. There are no longer any Unreviewed Concepts! See "MeSH Supplementary Concept Editing for 2002" below.

For the first time we include 67,427 names from MedDRA, the Medical Dictionary for Regulatory Activities Terminology (MedDRA 4.0) in the Metathesaurus.

The Veterans Health Administration National Drug File, VANDF, contributes 33,972 names; this new addition is included as the beginning of an effort to represent Clinical Drugs in a normal form within the Metathesaurus. See extended "VANDF and Clinical Drugs" below.

The Metathesaurus version of the Minimal Standard Terminology Digestive Endoscopy (MST) contributes English, French, and Italian Gastrointestinal Endoscopy terminology (see below).

A cumulative MRCUI file lets you find retired (merged or deleted) CUIs. This file will include more mappings for deleted CUIs in future editions. See Section 2.7.1.2.19.8 for more information.

** MRCOLS has an added field at the end of each row with the SQL92 data type for the column (See Section 2.7.1.2.2).

** There are two new relationships in MRREL:

RQ - related and possibly synonymous

SY - source asserted synonymy

These relationships are required to allow the Metathesaurus to correctly represent relationships asserted in its source vocabularies.

These new relationships allow us to represent Source Vocabularies transparently, and to label the source of each relationship with the asserting Source's Source Abbreviation (SAB). This means that users can distinguish Metathesaurus-created content from the sources', and to represent differing views. These views may well differ, for example SY relationships may be between strings in the same Concept (as expected) or between different Concepts (where the Metathesaurus disagrees).

MetamorphoSys

MetamorphoSys was updated for the 2002AA release. It includes many new features like being version aware so it knows which release it should be running against, the creation of source families to allow sources that should be removed together to be identified, ability to edit the entire precedence file, suppressible data can now be removed from subset created, and a progress monitor to track the process while running. For more information on the new features, go to Section 2.8 or read the tutorial given at the 2001 AMIA conference, available at http://umlsinfo.nlm.nih.gov.

We remind users to use MetamorphoSys, software which allows UMLS users to exclude any vocabulary which is not helpful in their applications or for which they have not negotiated the additional license arrangements required by the license agreement for their intended uses of the Metathesaurus. MetamorphoSys may also be used to create a customized version of the Metathesaurus files with added suppressibility and/or employing a user-defined order of precedence to determine the preferred name of any UMLS concept. This has been tested for Solaris, Windows XP, NT, Windows 98, Windows 2000, Windows ME, and Linux systems; all require 8 GB of free disk space.

MeSH Supplementary Concept Editing for 2002

The MeSH Supplementary Concept Records (SCR's, previously known as the Supplementary Chemical Records) have been included in the UMLS for many years.

With the change to a concept-oriented maintenance environment for MeSH, it became possible to edit the SCRs into a concept-based structure. Over the past year, almost all of the MeSH were edited as concepts and are no longer Unreviewed in the Metathesaurus.

Background: In previous editions because of the sheer numbers involved in re-editing these records, these records were released as is without re-editing and given the Concepts Attributes Status of Unreviewed (U). Semantic types were assigned algorithmically, based on the MeSH Main Headings to which the SCR was mapped to in MeSH. At present, the semantic types of the SCRs are still assigned algorithmically.

VANDF and Clinical Drugs

TTY	Type of term	count
CD	VA product name (clinical drug)	11345
HT	VA Classes Hierarchy	490
IN	VA Ingredient, Usually active	3962
SCD	Semantic Clinical Drug	7997
SCDC	Semantic Drug Component	10178

A portion of the VA National Drug Formulary (VANDF) is included for the first time in the 2002AA Metathesaurus. Based on ingredient information provided to NLM by the VA, the VANDF was used as the basis for creating a new concept in the Metathesaurus (see below), in a "Semantic Normal Form" (SNF). Not every drug in the VANDF could

be successfully parsed into the SNF Drug Component and the SNF Clinical Formulation. Accordingly, those concepts were not created. In future editions. more Semantic Normal Forms for Clinical Formulations and Drug Components will be created, and more mapping of clinical drugs will be accomplished.

We are creating two additional different types of concepts of semantic type "Clinical Drug" within the Metathesaurus.

The first type of concept within the Metathesaurus is that of the SNF Drug Component.

The first concept consists of the following:

> CUI|MetaID|ShortName|PreciseIngredient|ActiveIngredient|Basis|Strength |Units|Notes

The ingredients here are those named generically.

The second type of Clinical Drug concept within the Metathesaurus is that of the SNF Clinical Formulation.

It consists of the following:

> CUI|MetaID|ShortName|Component|ClinicalDoseForm|Notes

A ClinicalDoseForm is a Meta concept with source "Proposed HL7 Orderable Dose Forms." Each Component is a Metathesaurus Concept as defined above and has a relationship to this concept with attribute of "contains."

Every concept in the Metathesaurus of semantic type "Clinical Drug" will either be of these two types or related to one of these two types with the following relationship attributes:

> mapped_to (for clinical drugs with Manufactured Dose Form more specific than ClinicalDoseForm, or for clinical drugs with trade names)
>
> inverse_isa (for incompletely specified clinical drugs, e.g., Ventolin Repetabs)

Kits, Injectable Syringes, Applicators, and other devices used for administering drugs will be treated as Medical Devices.

Concepts for special devices which contain clinical drugs will be created as Drug Administering Devices and consist of the following:

> CUI|MetaID|ShortName|Device|ClinicalFormulation

where ClinicalFormulation is a Meta concept as defined above and has a relationship to this concept with attribute of "contains."

By convention, devices such as inhalers, patches, and other forms of controlling the administered amount of a drug will be treated as ClinicalDoseForms in the creation of ClinicalFormulations but not as DrugAdministeringDevices.

Metathesaurus version of the Minimal Standard Terminology Digestive Endoscopy: International Edition, April 22, 1998

```
Copyright 1995 European Society of Gastrointestinal Endoscopy (ESGE)

This is the Metathesaurus version, pending final organizational
approval.

(MTHMST2001, MTHMSTFRE, MTHMSTITA)

Here are the counts of terms:

 Language  SAB         Count
 --------  ---         -----
  English  MTHMST2001   1945
  French   MTHMSTFRE    1833
  Italian  MTHMSTITA    1799
```

0.1.2 Semantic Network

There were no changes from the 2001AA Semantic Network.

The total number of semantic types in the 2002AA edition is 134, the total number of relationships is 54.

0.1.3 SPECIALIST Lexicon and Lexical Programs

The Lexical Tools are now in Java. The Lexical tools distributed with the 2002 release are written in Java. Previously the tools were written in C. The motivations for this change are portability and maintenance. The Java based lexical tools can support more platforms than the C based versions. The lack of uniformity in C compilers, shells, "make" utilities etc. severely hampered installation of the lexical tools. The Java version avoids these pitfalls. Installation scripts will make installation faster and simpler. The Java based tools will be easier to maintain and develop. The C version will remain available and may be useful for high intensity tasks requiring greater speed than the Java version can deliver.

Prerequisites - The Java version of lexical tools will no longer require an ANSI C compiler, PERL or a "make" utility. Nor will it require the separate acquisition and installation of the Sleepycat Berkeley Btree software.

The only prerequisite now is a Java runtime executable (JRE) for versions of Java that are 1.2 or newer. The JRE for certain platforms is supplied with the Lexical Tools.

Other changes to the Lexical Tools:

1. New Command Line Interface

A cleaner, more consistent system of flow options has been developed that allows for the addition of new flow components and configurations. Single-letter flow options have been replaced with multi-letter options and explicit separators are used to attach arguments to flow options.

2. Changes in Output format for LVG

When inflecting terms LVG now identifies the uninflected form separately from the singular, infinitive or positive form. The need to differentiate these is apparent in fixed plural nouns like "police" which have no singular form.

3. A Change in the behavior of LuiNorm

LuiNorm now abstracts away form spelling variation as well as inflectional variation. For example: "color" and "colour" now Lui-normalize to "color".

Changes to the Specialist Lexicon:

1. Chemical terms removed.

Complex Structural chemical terms like "2'-(N,N-dimethylamino)-6-naphthoyl-4-cyclohexanecarboxylic acid" create problems for normalization and other lexical operations. An effort has been made to remove structural chemical terms from the lexicon through a process of automatic identification and human review. Not all chemical terms have been removed.

2. Growth

The 2002 release of the lexicon will include over 163,000 lexical entries covering 268,000 strings.

For more information, see:

http://umlslex.nlm.nih.gov/lvg/docs/userDoc/index.html

0.1.4 UMLS Knowledge Source Server

A new version of the Knowledge Source Server will be available in early 2002. For information on this new system, check:

http://umlsks.nlm.nih.gov

0.2 Introduction to the UMLS Documentation

Note: The latest and most authoritative documentation is always available at:

http://www.nlm.nih.gov/research/umls/UMLSDOC.HTML

Please consult this and other NLM Web resources for updates and additional information.

This documentation provides information about the Unified Medical Language System (UMLS) Knowledge Sources and related lexical programs. The Metathesaurus, the Semantic Network, and the SPECIALIST lexicon are products of the U.S. National Library of Medicine. These Knowledge Sources are designed primarily for use by system developers. They are meant to be consulted and used by applications programs to interpret and refine user queries, to map the user's terms to appropriate controlled vocabularies and classification schemes, to interpret natural language, and to assist in structured data creation. They are also useful as reference tools for database builders, librarians and other information professionals.

The documentation is arranged as follows:
What's New for the 2002AA UMLS (Section 0)
Introduction to the entire UMLS project (Section 1)
Metathesaurus (Section 2)
Semantic Network (Section 3)
SPECIALIST lexicon and related lexical software tools (Section 4)
Using the UMLS Knowledge Source Server via the Internet (Section 5)
Using the UMLS CD-ROMs (Section 6)

More detailed information concerning some items discussed in the manual sections above is included in the Appendices.

To obtain access to the UMLS Knowledge Sources, you must sign a license agreement. Information regarding the license agreement can be found in Appendix A. A copy of the license agreement is available with this documentation: 2001 License Agreement.

If there are questions on the use of the UMLS Knowledge Sources, the following resources are available:

UMLS users can subscribe to a listserver to discuss UMLS issues. To subscribe to the listserver, simply send a message to listserver@nlm.nih.gov which includes the following line:

subscribe umls-users <your full name>

To post a message to the umls-users listserver AFTER subscribing, send email to umls-users@lhc.nlm.nih.gov

Questions concerning the Metathesaurus can be sent to umlsmeta@nlm.nih.gov

Questions concerning the SPECIALIST lexicon can be sent to umlslex@nlm.nih.gov

Questions concerning the Knowledge Source Server can be sent to umlsks@nlm.nih.gov

A new website has been developed to assist UMLS users and potential UMLS users in gaining more information on the UMLS Knowledge Sources.

http://umlsinfo.nlm.nih.gov/

This site offers FAQs, other information, learning resources, and some examples. Use does not require a signed UMLS license agreement, since it does not provide access to information from vocabulary providers.

SECTION 1
OVERVIEW OF THE UMLS PROJECT

1.1 Purpose

The UMLS project is a long-term NLM research and development effort designed to facilitate the retrieval and integration of information from multiple machine-readable biomedical information sources. The sources of interest include: descriptions of the biomedical literature, clinical records, factual databanks, knowledge-based systems, and directories of people and organizations. Major barriers to effective retrieval and integration of information from these sources include (1) the variety of vocabularies and classifications used in different sources and by different users and (2) the sheer number and wide distribution of potentially relevant information sources. These barriers deter health professionals and researchers from using available machine-readable information and also hamper the development of effective search interfaces that might assist these users.

1.2 The Knowledge Sources Development Strategy

The UMLS approach involves the development of "intellectual middleware" in the form of machine-readable "Knowledge Sources" that can be used by a wide variety of applications programs to compensate for differences in the way concepts are expressed in different machine-readable sources and by different users. The goal is to make it easier to develop systems that link information from patient record systems, bibliographic databases, factual databases, expert systems, etc. The UMLS Knowledge Sources can also facilitate the development of data creation and indexing applications.

There are three UMLS Knowledge Sources, which are the following: The Metathesaurus (Section 2) contains semantic information about biomedical concepts, their various names, and the relationships among them. It is built from thesauri, classifications, coding systems, and lists of controlled terms that are developed and maintained by many different organizations. It contains and interconnects many standard clinical and biomedical vocabularies including those proposed as national standards under the administrative simplification provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The Semantic Network (Section 3) is a network of the general categories or semantic types to which all concepts in the Metathesaurus have been assigned. The SPECIALIST lexicon (Section 4) contains syntactic information about biomedical terms and will eventually cover the majority of component terms in the concept names present in the Metathesaurus. A number of lexical programs are distributed with the UMLS Knowledge Sources for use with the lexicon and the Metathesaurus; these are powerful tools for searching, indexing, and lexical processing.

The UMLS project is directed by a multi-disciplinary team of NLM staff and also involves research contractors. Apelon (the merger of Lexical Technology Inc. and Ontyx) has been instrumental in the development of the structure of the Metathesaurus

and the system used to build and maintain it. More than 1000 institutions and individuals around the world have licensed the UMLS Knowledge Sources and are applying them in a variety of applications and research projects. NLM itself makes use of the UMLS Metathesaurus in NLM Gateway and PubMed to enhance MEDLINE searching. The library also employs the Metathesaurus, the Semantic Network, the SPECIALIST lexicon, and the lexical programs in natural language processing and indexing research.

The strategy for the development of all UMLS components is to build successive approximations of the capabilities ultimately desired. The current edition of the UMLS Knowledge Sources supersedes the previous edition and includes substantial additional content.

1.3 Sources of Additional Information

Current Fact Sheets on the UMLS project, each of the three UMLS Knowledge Sources, the UMLS Knowledge Source Server, and NLM applications and research projects that make use of the UMLS Knowledge Sources are available from NLM's World Wide Web Server: http://www.nlm.nih.gov. The server also includes a comprehensive bibliography on the UMLS project covering papers published from 1986-96 and pointers to information about some of the UMLS applications developed by other institutions. Recent relevant articles can be retrieved in MEDLINE under the subject heading Unified Medical Language System.

A public website has been developed to assist UMLS users and potential UMLS users in gaining more information on the UMLS Knowledge Sources.

http://umlsinfo.nlm.nih.gov/

This site offers FAQs, other information, learning resources, and some examples. Use does not require a signed UMLS license agreement, since it does not provide access to information from vocabulary providers.

SECTION 2
METATHESAURUS

2.0 INTRODUCTION

The Metathesaurus is the central vocabulary component of the UMLS. The term Metathesaurus draws on Webster's Dictionary third definition for the prefix "Meta," i.e., "more comprehensive, transcending." In a sense, the Metathesaurus transcends the specific vocabularies and classifications it encompasses.

The Metathesaurus is a database of information on concepts that appear in one or more of a number of different controlled vocabularies and classifications used in the field of biomedicine. In general, the scope of the Metathesaurus is determined by the combined scope of its source vocabularies. **The Metathesaurus preserves the meanings, attributes, hierarchical connections, and other relationships between terms present in its source vocabularies, while adding certain basic information about each of its concepts and establishing synonymy and new relationships between concepts and terms from different source vocabularies**.

The Metathesaurus is used in a wide range of applications including: information retrieval from databases with human assigned subject index terms and from free-text information sources; linking patient records to related information in bibliographic, full-text, or factual databases; natural language processing and automated indexing research; and structured data entry. In many cases, the utility of the Metathesaurus is enhanced when it is used in combination with the SPECIALIST lexicon, the lexical programs, and the Semantic Network. To obtain coherent, comparable results in data creation applications, such as patient data entry, it is necessary to define which Metathesaurus concepts and terms can be included in the records being created. This may be done by selecting one or more of the many Metathesaurus source vocabularies which provide the most appropriate concepts and terms for the specific data being created. Other Metathesaurus concepts and terms will then provide synonyms and related terms which can help to lead users to the vocabulary selected for a particular data creation application.

2.1 SOURCE VOCABULARIES

The Metathesaurus contains concepts and concept names from more than 60 vocabularies and classifications, some in multiple editions. (A complete list of the current Metathesaurus source vocabularies appears in Appendix B, Section B.2.). Many of the source vocabularies are included in their entirety; for others the Metathesaurus has partial coverage. (The number of concept names from each source is detailed in Appendix B, Section B.3). Some material in the UMLS Metathesaurus is from copyrighted sources of the respective copyright claimants.

The Metathesaurus source vocabularies include terminologies designed for use in patient-record systems; large disease and procedure classifications used for statistical reporting and billing; more narrowly focused vocabularies used to record data related to psychiatry,

nursing, medical devices, adverse drug reactions, etc.; disease and finding terminologies from expert diagnostic systems, and some thesauri used in information retrieval.

The Metathesaurus structure can incorporate translations of its source vocabularies into languages other than English. The 2002AA Metathesaurus includes the Dutch, French, Finnish, German, Italian, Portuguese, Russian (transliterated), and Spanish translations of NLM's Medical Subject Headings (MeSH). This edition also includes German translations of ICD10 and UMDNS; ICPC terms in Basque, Danish, Dutch, Finnish, French, German, Hebrew, Hungarian, Italian, Norwegian, Spanish, and Swedish; CPT terms in Spanish; Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy in French and Italian; and WHOART terms in French, German, Portuguese, and Spanish.

Users should determine which vocabularies are useful for their purposes, and which vocabularies would require additional license arrangements for use. MetamorphoSys (see Section 2.8) should be used to exclude vocabularies; it removes all vocabulary information and thus ensures compliance with the UMLS License Agreement.

2.2 ORGANIZATION BY CONCEPT

The Metathesaurus is organized by concept or meaning. In essence, its purpose is to link alternative names and views of the same concept together and to identify useful relationships between different concepts. Each concept or meaning in the Metathesaurus has a unique concept identifier (CUI) which itself has no intrinsic meaning. Each unique concept name or string in each language in the Metathesaurus has a unique string identifier (SUI). Any variation in upper-lower case is a separate string, with a separate SUI. The same string in different languages (e.g., English and Spanish) will have a different string identifier for each language. For English language entries in the Metathesaurus only, each string is linked to all of its lexical variants or minor variations by means of a common term identifier (LUI). (In the Metathesaurus, therefore, an English "term" is the group of all strings that are lexical variants of each other.) All string and term identifiers are linked to at least one concept identifier. Different terms with the same meaning are linked to the same concept identifier. Thus, in the Metathesaurus, strings are linked to terms and both strings and terms are linked to concepts.

FIGURE 1.

<table>
<tr><th>Concept (CUI)</th><th>Terms (LUIs)</th><th>Strings (SUIs)</th></tr>
<tr><td rowspan="4">C0004238
Atrial Fibrillation
(preferred)
Atrial Fibrillations
Auricular Fibrillation
Auricular Fibrillations</td><td rowspan="2">L0004238
Atrial Fibrillation
(preferred)
Atrial Fibrillations</td><td>S0016668
Atrial Fibrillation
(preferred)</td></tr>
<tr><td>S0016669
Atrial Fibrillations</td></tr>
<tr><td rowspan="2">L0004327
(synonym)
Auricular Fibrillation
Auricular Fibrillations</td><td>S0016899
Auricular Fibrillation
(preferred)</td></tr>
<tr><td>S0016900
(plural variant)
Auricular Fibrillations</td></tr>
</table>

For example, in Figure 1, the string "Atrial Fibrillation" and its plural "Atrial Fibrillations" have different string identifiers, but are linked to the same term identifier. "Auricular Fibrillation" and its plural "Auricular Fibrillations" are linked to a different term identifier. Since "Atrial Fibrillation" and "Auricular Fibrillation" have been judged to have the same meaning, their different term identifiers are linked to the same concept identifier. Note that users may select preferred concept names by source; see Section 2.8.

2.2.1 ORDER OF PRECEDENCE AMONG CONCEPT NAMES

Users of the Metathesaurus should configure their applications to feature terminology from the source vocabularies most appropriate to their user populations. For example, concept names from SNOMED International might be preferred in some clinical applications, and terminology from MeSH might be preferred in some literature retrieval systems. The source information provided in the Metathesaurus allows different applications to pick different preferred names for the same concepts; MetamorphoSys (Section 2.8) can be used to make this change.

As a convenience for those who build the Metathesaurus, one string from each term is designated as the default preferred name in the Metathesaurus. For example, in Figure 1: -"Atrial Fibrillation" is labeled as the preferred form of its term; -"Atrial Fibrillations" is labeled as a plural variant of this preferred term; -"Auricular Fibrillation" is labeled as the preferred form of its term; -"Auricular Fibrillations" is labeled as a plural variant of this preferred form. One of the terms is designated as the preferred name of the concept as a whole. In this case, the "Atrial Fibrillation" term is designated as the preferred term and therefore its preferred form, "Atrial Fibrillation," is the default preferred name of the concept in the Metathesaurus.

To avoid laborious selection among alternative terms and forms of terms, selection of the default preferred name for any Metathesaurus concept is based on an order of precedence

of Metathesaurus source vocabularies. Different types of strings, e.g., preferred terms, cross references, abbreviations, from the same vocabulary will have different positions in this order. The factors considered in establishing the order include breadth of subject coverage, frequency of update, and the degree to which the source's concept names are used in regular clinical or biomedical discourse. The order of precedence used appears in Appendix B, Section B.5

2.2.2 STRINGS WITH MULTIPLE MEANINGS

In a small number of cases, the same string (ignoring differences in upper-lower case) may name different concepts, usually in different Metathesaurus source vocabularies. In the abbreviated example that follows, the string "Cold" is a name for the temperature in one vocabulary. In another vocabulary, "Cold" is an alternate name for the "Common cold". In a third Vocabulary, "COLD" is an acronym for "chronic obstructive lung disease". As a result, "Cold" or "COLD" appears as a name of at least three different concepts in the Metathesaurus. As illustrated in the sample in Figure 2, three distinct strings, "Cold <1>", "Cold <2>" and "COLD<3>" have been created for the Metathesaurus to give each meaning a unique name. Each of these strings is linked to only one concept. The plain strings "Cold" and "COLD" have explicit "ambiguous string" indicators in the Metathesaurus. More descriptive names have also been created by Metathesaurus editors to distinguish some of the strings with multiple meanings, e.g., "cold temperature". These names are identified by the term group "PN".

FIGURE 2.

Concepts (CUIs)	Terms (LUIs)	Strings (SUIs)
C0009264 **cold temperature**	**L0215040** cold temperature	**S0288775** cold temperature
	L0009264 Cold <1> Cold	**S0007170** Cold <1>
		S0026353 Cold
C0009443 **Common Cold**	**L0009443** Common Cold	**S0026747** Common Cold
	L0009264 Cold <2> Cold	**S0007171** Cold <2>
		S0026353 Cold

C0024117 **Chronic Obstructive Airway Disease**	**L0498186** Chronic Obstructive Airway Disease	**S0837575** Chronic Obstructive Airway Disease
	L0008703 Chronic Obstructive Lung Disease	**S0837576** Chronic Obstructive Lung Disease
	L0009264 COLD <3> COLD	**S0829315** COLD <3>
		S0474508 COLD

2.3 RELATIONSHIPS BETWEEN DIFFERENT CONCEPTS

The Metathesaurus also represents several types of relationships between different concepts. Many relationships are derived directly from the source vocabularies. For example, the fact that there is a relationship between "Atrial Fibrillation" and "Arrhythmia" is derived from the hierarchical tree structures in the Medical Subject Headings (MeSH). The exact nature of the relationship, i.e., "Atrial Fibrillation" is_a "Arrhythmia", is not yet present in MeSH, but was added during the construction of the Metathesaurus. Relationships between concepts from different source vocabularies were also created during Metathesaurus construction. For example, the COSTAR concept "Paroxysmal Atrial Fibrillation" is identified in the Metathesaurus as having a narrower-than relationship to "Atrial Fibrillation", a concept that is present in MeSH, ICD-9-CM, SNOMED, and COSTAR, etc.

2.3.1 RELATIONSHIPS IN THE METATHESAURUS

Relationships in the Metathesaurus come from the sources themselves or are created by the Metathesaurus editors to link concepts that would not otherwise be connected.

There are eleven types of relationships that exist in the Metathesaurus.

Broader (RB)	has a broader relationship.
Narrower (RN)	has a narrower relationship.
Other related (RO)	has relationship other than synonymous, narrower, or broader.
Like (RL)	the two concepts are similar or "alike". In the current edition of the

	Metathesaurus, most relationships with this attribute are mappings provided by a source, named in SAB and SL; hence concepts linked by this relationship may be synonymous, i.e. self-referential: CUI1 = CUI2. In previous releases, some MeSH Supplementary Concept relationships were represented in this way.
RQ	related and possibly synonymous.
SY	source asserted synonymy.
Parent (PAR)	has parent relationship in a Metathesaurus source vocabulary.
Child (CHD)	has child relationship in a Metathesaurus source vocabulary.
Sibling (SIB)	has sibling relationship in a Metathesaurus source vocabulary.
AQ	is an allowed qualifier for a concept in a Metathesaurus source vocabulary.
QB	can be qualified by a concept in a Metathesaurus source vocabulary.

2.3.2 CONTEXTS IN THE METATHESAURUS

Starting with the 2002AA release, sources with contexts now have "Full" contexts, i.e., all levels of terms may have Ancestors, Parents, Children and Siblings. Title and mini contexts have been eliminated. A Full context may also be further designated to be one of Multiple, Nosib (No siblings) or both Multiple and Nosib.

> "Multiple" indicates that a single concept in this source may have multiple hierarchical positions.
>
> No siblings (Nosib) indicates that siblings have not been computed for this source.

Listing of sources with contexts (by source abbreviation):

AIR93:FULL-NOSIB-MULTIPLE
ALT2000:FULL
AOD99:FULL
CCS99:FULL
CPT2002:FULL-NOSIB
CSP2001:FULL-MULTIPLE
CST95:FULL-MULTIPLE
DSM3R:FULL-NOSIB
DSM4:FULL-NOSIB
HCPCS02:FULL-NOSIB
HHC96:FULL

HL7:FULL-MULTIPLE
ICD10AM:FULL
ICD10:FULL-NOSIB
ICD2002:FULL
ICPC2E:FULL-NOSIB-MULTIPLE
ICPC2P:FULL-NOSIB-MULTIPLE
ICPC93:FULL-NOSIB-MULTIPLE
MDR40:FULL-MULTIPLE
MSH2002:FULL-MULTIPLE
NAN99:FULL
NCI2001a:FULL-MULTIPLE
NEU99:FULL
NIC99:FULL-NOSIB-MULTIPLE
NOC97:FULL
OMS94:FULL-MULTIPLE
PCDS97:FULL
PDQ2001:FULL-MULTIPLE
PPAC98:FULL
PSY2001:FULL-NOSIB-MULTIPLE
RCD99:FULL-MULTIPLE
SNM2:FULL-NOSIB-MULTIPLE
SNMI98:FULL-NOSIB
UMD2002:FULL-MULTIPLE
UWDA155:FULL-MULTIPLE
VANDF01:FULL-NOSIB-MULTIPLE
WHO97:FULL-MULTIPLE

2.4 CONCEPT AND STRING ATTRIBUTES

All other pieces of information in the Metathesaurus are either attributes of a concept or attributes of a string. For example, the semantic types "Pathologic Function" and "Finding" and the definition "Disorder of cardiac rhythm characterized by rapid, irregular atrial impulses and ineffective atrial contractions." are attributes of the concept with the preferred name "Atrial Fibrillation". The various sources of the string "Atrial Fibrillation" e.g., SNM2|PT|F73570, ICD2002|PT|427.31, are string attributes.

The structure of the Metathesaurus distribution files facilitates access to concept relationships and to string and concept attributes, irrespective of the string used as a point of entry to a concept.

2.5 DATA ELEMENTS

2.5.0 INTRODUCTION

The Metathesaurus is built to be concept oriented; and there is one key identifier, the Concept Unique Identifier or CUI, which is used to link all information about a concept.

CUIs connect all names for a concept, are used to link concepts in relationships, and identify all attributes of a concept.

There are also Lexical Unique Identifiers (LUIs) which identify and link all concept names which are lexical variants of each other; and String Unique Identifiers (SUIs) which identify and link all identical strings.

The Metathesaurus has four logical groups of data elements, which are briefly described in the following sections. A complete description of the data elements that make up the Metathesaurus files appears in Appendix B, Sections B.1.1 (Column Descriptions) and B.1.2 (Attribute Descriptions).

2.5.1 CONCEPT NAMES

In addition to the Unique Identifiers discussed above, other columns specify certain characteristics of each concept. The concept name columns are as follows:

CUI	Unique Identifier for Concept
LAT	Language of Term
LRL	Least Restriction Level
LUI	Unique Identifier for Term
STR	String
STT	String Type
SUI	Unique Identifier for String
TS	Term Status

See Section B.1.1 for descriptions of each column and the file MRCON.

2.5.2 RELATIONSHIPS

These data elements describe relationships between different Metathesaurus concepts or meanings. By definition, these are non-synonymous relationships, although the combination of concepts represented in an Associated Expression may have a meaning equivalent to the single concept to which the expression is linked.

Relationships may be derived from the Metathesaurus source vocabularies or be added expressly for the Metathesaurus. The relationship data elements are as follows:

ATX	Associated Expressions
COC	Co-occurring Concepts
REL	Related Concepts

See Section B.1.1 for descriptions of these columns and the files MRREL, MRCOC and MRATX.

2.5.3 ATTRIBUTES

These data elements provide additional information that helps to define the meaning of a concept and explain how it has been used in various information sources; identify special kinds of terms; or provide information about the source(s) of the unique strings in the Metathesaurus. There are more than 110 attributes in the Metathesaurus. A few examples are listed below. See Appendix B, Section B.1.2 for a complete list.

Concept Attributes	
DEF	Definition
RN	Registry Number
STY	Semantic Type

See Section B.1.2 for descriptions of all the attributes and the file MRSAT.

2.5.4 VOCABULARY SOURCE INFORMATION AND CONTEXTS

Each occurrence of a string (name) in a source is identified by:

SAB	Source abbreviation
SCD	Code or identifier in source vocabulary
TTY	Term Type

See MRSO file for more information.

The hierarchies in each source are represented in the MRCXT file; distance-one hierarchical relationships are also found in the MRREL file. See MRCXT and MRREL files.

2.6 CONCEPT NAME INDEXES

2.6.0 INTRODUCTION

To assist system developers in building applications that retrieve all strings or concept names which include specific words or groups of words, three indexes to the concept names are provided: a Word Index, a Normalized Word Index (for English words only), and a Normalized String Index (for English strings only). The indexes are described in sections 2.6.1, 2.6.2, and 2.6.3 respectively. To make the distinctions among them clearer, the examples include words or strings that would appear in each index for the

following set of Metathesaurus concept names:

Lung Diseases, Obstructive	(C0024117, L0024117, S0058463)
Obstructive Lung Diseases	(C0024117, L0024117, S0068169)
Lung Disease, Obstructive	(C0024117, L0024117, S0058458)
Obstructive Lung Disease	(C0024117, L0024117, S0068168)

2.6.1 WORD INDEX

2.6.1.1 Description

The word index connects each individual word in any Metathesaurus string to all its related string, term, and concept identifiers. There are separate word index files for each language in the Metathesaurus.

There is one entry for each word found in each unique string in each language. Each entry has five subelements.

1. LAT
2. WD
3. CUI
4. LUI
5. SUI

2.6.1.2 Definition of a Word

For the purpose of creating this index, a word is defined as a token containing only alphanumeric characters with length one or greater; for more information, see the SPECIALIST Lexicon and tools.

2.6.1.3 Word Index Example

For the four concept names listed in Section 2.6.0, the word index will contain multiple entries for each of the following words: disease, diseases, lung, obstructive. Two of the entries generated for the names "Lung Disease, Obstructive" and Obstructive Lung Disease" are shown below:

ENG|disease|C0024117|L0024117|S0058458|

ENG|disease|C0024117|L0024117|S0068168|

2.6.2 NORMALIZED WORD INDEX

2.6.2.1 Description

The normalized word index connects each individual normalized English word to all its related string, term, and concept identifiers.

There is one entry for each normalized word found in each unique English string. There are no entries for other languages in this index. Each entry has five subelements.

1. LAT (always ENG in this edition of the Metathesaurus)
2. WD
3. CUI
4. LUI
5. SUI

2.6.2.2 Definition of Normalized Word

The normalization process involves breaking a string into its constituent words, lowercasing each word and converting it to its uninflected form. Normalized words are generated by uninflecting each word and stripping out a small number of stop words. The uninflected forms are generated using the SPECIALIST lexicon if the words appear in the lexicon; otherwise they are generated algorithmically.

2.6.2.3 Normalized Word Example

For the four concept names listed in Section 2.6.0 the normalized word index will contain multiple entries for each of the following words: disease, lung, obstructive. Since the normalized word index contains base forms only, it does not contain entries for the plural "diseases". In this index, therefore, all four concept names are linked to the normalized word "disease", as follows:

ENG|disease|C0024117|L0024117|S0058458|

ENG|disease|C0024117|L0024117|S0058463|

ENG|disease|C0024117|L0024117|S0068168|

ENG|disease|C0024117|L0024117|S0068169|

2.6.3 NORMALIZED STRING INDEX

2.6.3.1 Description

The normalized string index connects the normalized form of a Metathesaurus string to all its related string, term, and concept identifiers. There is one entry for each unique (non-normalized) English string. There are no entries for other languages in this index. Each entry has five subelements.

1. LAT (always ENG in this edition of the Metathesaurus)
2. NSTR
3. CUI
4. LUI
5. SUI

2.6.3.2 Definition of Normalized String

The normalization process involves breaking a string into its constituent words, lowercasing each word, converting each word to its uninflected form, and sorting the words in alphabetic order. Normalized strings are generated by uninflecting each word leaving out a small number of stop words. The uninflected forms are generated using the SPECIALIST lexicon if the words appear in the lexicon; otherwise they are generated algorithmically.

2.6.3.3 Normalized String Example

Since the four concept names listed in Section 2.6.0 are composed of the same set of normalized words, the Normalized String Index will contain four entries for a single string: disease lung obstructive, in which the component normalized words appear in alphabetical order. The complete set of Normalized String Index entries generated by the four concept names is as follows:

ENG|disease lung obstructive|C0024117|L0024117|S0058458|

ENG|disease lung obstructive|C0024117|L0024117|S0058463|

ENG|disease lung obstructive|C0024117|L0024115|S0068168|

ENG|disease lung obstructive|C0024117|L0024117|S0068169|

2.6.4 WORD INDEX PROGRAMS

The programs that generate these indexes are written in Java. They may be of use to system developers who are developing their own interfaces to the UMLS data or for other

purposes. Section 4 includes information about these and other lexical programs provided with the UMLS Knowledge Sources.

2.7 FILE FORMAT

2.7.0 INTRODUCTION

The Metathesaurus data is distributed in an ASCII Relational Format.

2.7.1 ASCII RELATIONAL FORMAT

2.7.1.1 General Description

The Metathesaurus relational format is not fully normalized. By design, there is duplication of data among different relations and within certain relations. In particular, links between different Metathesaurus concepts appear twice (e.g., from entry A to entry B and from entry B to entry A). Developers will need to make their own decisions about the extent to which this redundancy should be retained, reduced, or increased for their specific applications.

All file names in the relational format begin with the letters MR.

All files except MRRANK are sorted by row.

2.7.1.1.1 Data Files

The data in each Metathesaurus entry may be represented in more than 20 different "relations" or files. These files correspond to the four logical groups of data elements described in Section 2.5 and the indexes described in Section 2.6 as follows:

Metathesaurus Concept Names (2.5.1) = MRCON
Relationships between Different Concept Names (2.5.2) = MRREL, MRCOC, MRATX
Attributes (2.5.3) = MRSAT, MRDEF, MRSTY, MRLO, MRRANK
Source Information and contexts (2.5.4) = MRSO, MRCXT
Indexes (2.6) = MRXW.BAQ, MRXW.DAN, MRXW.DUT, MRXW.ENG, MRXW.FIN, MRXW.FRE, MRXW.GER, MRXW.HEB, MRXW.HUN, MRXW.ITA, MRXW.NOR, MRXW.POR, MRXW.RUS, MRXW.SPA, MRXW.SWE, MRXNW.ENG, MRXNS.ENG

2.7.1.1.2 Ancillary Files

There are 5 ancillary files:

MRFILES	describes the data files.
MRCOLS	describes the columns or data elements in the data files.

AMBIG.LUI	contains all "Ambiguous" term identifiers in the Metathesaurus.
AMBIG.SUI	contains all "Ambiguous" string identifiers in the Metathesaurus.
MRCUI	contains all retired CUIs, with mappings where possible.

The AMBIG* files provide a convenient way to identify all Metathesaurus terms and strings that have more than one meaning in Metathesaurus source vocabularies.

2.7.1.1.3 Columns and Rows

Each relation or named table of data values has by definition a fixed number of columns; the number of rows depends on the content of a particular version of the Metathesaurus.

A column is a sequence of all the values in a given data element or logical subelement. In general, columns for longer variable length data elements will appear to the right of columns for shorter and/or fixed length data elements. The information for all columns in the relational files is described in Appendix B.1.1, Metathesaurus Column Descriptions.

A row contains the values for one or more data elements or logical subelements for one Metathesaurus entry. Depending on the nature of the data elements involved, each Metathesaurus entry may have one or more rows in a given file. The values for the different data elements or logical subelements represented in the row are separated by vertical bars (|). If an optional element is blank, the vertical bars are still used to maintain the correct positioning of the subsequent elements. Each row is terminated by a vertical bar and line termination.

2.7.1.2 Descriptions of Each File

2.7.1.2.1 Relation Relation (File = MRFILES)

There is exactly one row in this file for each physical segment of the files in the relational format. The columns or data elements in the file are:

Col.	Description
FIL	Physical FILENAME
DES	Descriptive Name
FMT	Comma separated list of COL, in order
CLS	# of COLUMNS
RWS	# of ROWS
BTS	Size in Bytes with <CR><LF> line termination

Sample Records

MRATX|Associated Expressions|CUI,SAB,REL,ATX|4|7295|442571|
MRCOC|Co-occurring Concepts|CUI1,CUI2,SAB,COT,COF,COA|6|9061980|343331578|
MRCOLS|Attribute Relation|COL,DES,REF,MIN,AV,MAX,FIL, DTY|8|115|5728|

2.7.1.2.2 Attribute Relation (File = MRCOLS)

There is exactly one row in this file for each column or data element in each file in the relational format.

Col.	Description
COL	Column or data element name
DES	Descriptive Name
REF	Documentation Section Number
MIN	Minimum Length, Characters
AV	Average Length
MAX	Maximum Length, Characters
FIL	Physical FILENAME in which this field occurs
DTY	Suggested SQL-92 data type for this column

Sample Records

ATN|Attribute name||2|3.15|7|MRSAT|varchar(20)|
ATV|Attribute value||1|9.71|3634|MRSAT|varchar(4000)|
ATX|Associated expression||5|35.89|242|MRATX|varchar(300)|

2.7.1.2.3 Concept Names (File = MRCON)

There is exactly one row in this file for each meaning of each unique string in the Metathesaurus, i.e., there is exactly one row for each unique CUI-SUI combination in the Metathesaurus. Any difference in upper-lower case, word order, etc. creates a different unique string.

Col.	Description
CUI	Unique identifier for concept
LAT	Language of Term
TS	Term status
LUI	Unique identifier for term

STT	String type
SUI	Unique identifier for string
STR	String
LRL	Least Restriction Level

Sample Records

C0002871|ENG|P|L0002871|PF|S0013742|Anemia|0|
C0002871|ENG|P|L0002871|VP|S0013787|Anemias|0|
C0002871|ENG|P|L0002871|VC|S0352787|ANEMIA|0|
C0002871|ENG|P|L0002871|VC|S0414880|anemia|0|
C0002871|ENG|P|L0002871|VO|S0470197|Anemia, NOS|3|
C0002871|ENG|S|L0280031|PF|S0803242|Anaemia|3|

2.7.1.2.4 Related Concepts (File = MRREL)

There is one row in this table for each relationship between Metathesaurus concepts known to the Metathesaurus, with the following exceptions found in other files: co-occurrences found in MRCOC; Locator information in MRLO; and Associated Expressions found in MRATX.

Note that for asymmetrical relationships there is one row for each direction of the relationship. Note also the direction of REL - the relationship which the SECOND concept (with Concept Unique Identifier CUI2) HAS TO the FIRST concept (with Concept Unique Identifier CUI1).

Col.	Description
CUI1	Unique identifier of first concept
REL	Relationship of second to first concept
CUI2	Unique identifier of second concept
RELA	Relationship attribute
SAB	Abbreviation of the source of relationship
SL	Source of relationship labels
MG	Machine-generated and unverified indicator (optional)

Sample Records

```
C0002871|CHD|C0002891|isa|MSH2002|MSH2002||
     Anemia, Neonatal (C0002891)
       has CHILD REL and isa RELA
         to Anemia (C0002871)
```

```
C0002871|RB|C0221016||MTH|MTH||
  Red blood cell disorder, NOS (C0221016)
    has broader REL
      to Anemia (C0002871)

C0002871|RL|C0002886|mapped_to|SNMI98|SNMI98||
  Anemia, Macrocytic (C0002886)
    has like relationship
     to Anemia (C0002871)

C0002871|RO|C0002886|clinically_associated_with|CCPSS99|CCPSS99||
  Megaloblastic anemia due to folate deficiency, NOS (C0151482)
    has clinically_associated_with relationship
      to Anemia (C0002871)
```

2.7.1.2.5 Co-occurring Concepts (File = MRCOC)

There are two rows in this table for each pair of concepts that co-occur in each information source represented one for each direction of the relationship. (Note that the COA data may be different for each direction of the relationship). Many Metathesaurus concepts have no entries in this file. Due to the very large number of co-occurrence relationships, they are distributed in a separate file.

Col.	Description
CUI1	Unique identifier of first concept
CUI2	Unique identifier of second concept Note: Where COT is MeSH topical qualifier (LQ) and CUI2 is not present, the count of citations of CUI1 with no MeSH qualifiers is reported.
SOC	Abbreviation of the Source of co-occurrence information if applicable
COT	Type of co-occurrence
COF	Frequency of co-occurrence, if applicable
COA	Attributes of co-occurrence, if applicable

Sample Records

```
C0002871||MED02|LQ|1||
C0002871|C0000530|MBD02|L|2|CI=1,EN=1,ME=1,PA=1|
C0002871|C0000727|MBD02|L|1|BL=1,ET=1|
C0002871|C0000737|MBD02|L|1|ET=1|
C0002871|C0000772|MBD02|L|2|CN=2|
```

Illustrative Examples (may not be current):

```
C0000039|C0001977|MBD00|L|5|CH=2,<>=1,AN=1,ME=1|
```

has frequency of five, two with CH subheading and one each for no subheading, AN, and ME. This means four co-occurring citations had subheadings (one had none) and since there are four subheading counts (2+1+1), these four had one SH each. But:

C0000039|C0023828|MBD00|L|31|CH=24,<>=5,ME=4,PD=2,PK=1,RE=1|

has frequency of 31, with 37 "Attribute" counts; five citations have no SH, while 26 have a total of 32 SHs - some citations had more than one SH.

In this example there is no CUI2:

C0005851||MBD00|LQ|6||

because there are 6 cases in MBD00 of citations with no MeSH qualifier

And finally, in this example, the co-occurrence is a positive one in a Knowledge Base, AI Rheum:

C0015967|C0026691|AIR93|KP|||

Co-occurrences are concepts that occur together in the same "entries" in some information source. The relationships represented here are obtained from machine-manipulation of the information source. Co-occurrence relationships may exist between similar concepts (e.g., "Atrial Fibrillation" and "Arrhythmia") or between very different concepts that nevertheless have some important connection in the field of biomedicine (e.g., "Atrial Fibrillation" and "Digoxin"), or between a primary concept and a qualifier e.g., "Lithotripsy" and "instrumentation". A co-occurrence relationship can exist between two concepts that have no other apparent relationship, although the frequency of such co-occurrences will be small.

In the current Metathesaurus, there are three sources of co-occurrence data: MEDLINE, AI/RHEUM, and CCPSS. From MEDLINE, co-occurrence data was computed for concepts that were designated as principal or main points in the same journal article i.e., the co-occurrence counts do not include articles in which either or both of the concepts were present and indexed in MEDLINE but not designated as main points. (A concept is considered to be a main point if the * is attached to the main heading or any of its subheadings.)

Two overall frequencies of MEDLINE co-occurrence are provided: one for recent MEDLINE data (MED??) and one for MEDLINE data from a preceding block of years (MBD??); see SOC for date ranges in the current edition. Separate counts are provided for the frequencies with which the first concept was qualified by different MeSH qualifiers or by no qualifier at all when it co-occurred with the second concept. There are separate entries for each direction of the co-occurrence relationship. The related subheading occurrence information in each entry belongs to the first concept in the entry and is therefore different for each direction of the relationship.

In addition to the specific qualifier information associated with two co-occurring concepts, this element also includes in entries with LQ and LQB values for type of co-occurrence, totals for the number of times each main concept was qualified by a specific subheading or by no subheading.

The AI/RHEUM co-occurrence data represent the co-occurrence of diseases and findings in the AI/RHEUM knowledge base, i.e., the diseases that co-occur with a particular finding and the findings that co-occur with a particular disease. Each disease/finding pair can co-occur only once in the AI/RHEUM knowledge base.

In CCPSS, the co-occurrence data is extracted from patient records and includes problem-problem co-occurrences within a patient record as well as problem-modifier co-occurrences.

2.7.1.2.6 Simple Concept, Term, and String Attributes (File = MRSAT)

There is exactly one row in this table for each concept, term and string attribute that does not have a sub-element structure. All Metathesaurus concepts have entries in this file.

Col.	Description
CUI	Unique identifier for concept
LUI	Unique identifier for term (optional)
SUI	Unique identifier for string (optional)
SCD	Unique identifier or code for entry in the source of the attribute, e.g., for all attributes derived from MeSH, the MeSH unique identifier (optional).
ATN	Attribute name. Possible values are all described in Appendix B, Section B.1.2.
SAB	Abbreviation of the source of the attribute. Allowed values are listed in Appendix B, Section B.2.)
ATV	Attribute value described under specific attribute name in Appendix B, Section B.1.2. A few attribute values exceed 1,000 characters.

Sample Records

```
C0002871|L0002871|S0013742|D000740|MMR|MSH2002|19960610|
C0002871|L0002871|S0013742|D000740|MN|MSH2002|C15.378.71|
C0002871|L0002871|S0013742|D000740|TH|MSH2002|POPLINE (1994)|
C0002871|L0002871|S0414880|208/04453|SOS|PDQ2001|secondary related condition|
C0002871|L0002871|S0470197|DC-10010|SIC|SNMI98|285.9|
```

2.7.1.2.7 Definitions (File = MRDEF)

There is exactly one row in this file for each definition in the Metathesaurus. A few definitions approach 3,000 characters in length.

Col.	Description
CUI	Unique identifier for concept
SAB	Abbreviation of the source of the definition
DEF	Definition

Sample Records

C0002871|MSH2002|A reduction in the number of circulating erythrocytes or in the quantity of hemoglobin.|

2.7.1.2.8 Sources (File = MRSO)

The vocabulary source(s) for a concept, term, and string.

There is exactly one row in this file for each source of each string in the Metathesaurus. All Metathesaurus concepts have entries in this file.

Col.	Description
CUI	Unique identifier for concept
LUI	Unique identifier for term
SUI	Unique identifier for string
SAB	Source abbreviation. Allowed values are listed in Appendix B, Section B.2
TTY	Term type in that source. Allowed values are listed in Appendix B., Section B.4.
SCD	Unique Identifier or code for string in that source.
SRL	Source Restriction Level

Sample Records

C0002871|L0002871|S0013742|CCS99|MD|4.1|0|
C0002871|L0002871|S0013742|ICPCPAE|PT|B82005|3|
C0002871|L0002871|S0013742|LCH90|PT|U000235|0|
C0002871|L0002871|S0013742|MSH2002|MH|D000740|0|
C0002871|L0002871|S0013742|MTH|PT|U000161|0|
C0002871|L0002871|S0013742|MTH|PT|U000164|0|
C0002871|L0002871|S0013742|PSY2001|PT|02450|3|
C0002871|L0002871|S0013742|RCDAE|PT|XM05A|3|

The information in MRSO can be used to determine whether a particular concept, name, or code is present in a particular source, and in what form it appears. This information can be used to extract from the Metathesaurus the terms appropriate for searching information sources that have been indexed by specific vocabularies or classifications. In the case of the COSTART vocabulary, the SCD Unique Identifier is actually an official abbreviation composed of alphabetic characters which in some cases may include a space.

NLM recommends that MetamorphoSys (described in Section 2.8) be used to exclude vocabularies because it reliably removes ALL source information.

2.7.1.2.9 Semantic Types (File = MRSTY)

There is exactly one row in this file for each semantic type assigned to each concept. All Metathesaurus concepts have at least one entry in this file. Many have more than one entry.

Col.	Description
CUI	Unique identifier of concept
TUI	Unique identifier of Semantic type
STY	Semantic type. The valid values are defined in the Semantic Network.

Sample Record

C0002871|T047|Disease or Syndrome|

2.7.1.2.10 Concept contexts (File = MRCXT)

There are rows in this file for each occurrence of a concept in a hierarchy in any of the UMLS source vocabularies - a "context" in this discussion. Many Metathesaurus concepts have multiple contexts while others may have none. The number of rows per context differs depending on the number of ancestor, sibling, or child terms the concept has in that context. Because some concepts have multiple contexts in the same source (e.g., MeSH), a context number (CXN - e.g., 1, 2, 3) is used to identify all members of the same context. The CXNs are not global but are created as required for each concept. Since some concepts have multiple contexts in the same vocabulary with the same SUI, each distinct context can be retrieved with a CUI-SUI-SAB-CXN key. The "distance-1 relationships," i.e., the immediate parent, immediate child, and sibling relationships, represented in this file are also present in the MRREL file.

(Note: The RELA was incorrectly called REL in versions before 2001.)

Col.	Description
CUI	Unique identifier of concept

SUI	Unique identifier for string used in this context
SAB	Source abbreviation. Allowed values are listed in Appendix B, Section B.2
SCD	Unique Identifier or code for string in that source.
CXN	The context number (to distinguish multiple contexts in the same source with the same SUI).
CXL	Context member label, i.e., ANC for ancestor of this concept, CCP for concept, SIB for sibling of this concept, CHD for child of this concept.
RNK	For rows with a CXL value of ANC, the rank of the ancestors (e.g., a value of 1 denotes the most remote ancestor in the hierarchy)
CXS	String for context member.
CUI2	Unique concept identifier of context member (may be empty if context member is not yet in the Metathesaurus).
HCD	Hierarchical number or code of context member in this source (optional).
RELA	Relationship attribute providing further categorization of the CXL, if applicable and known. Valid values are the Semantic Network Relationships (see the UMLS Semantic Network documentation in Section 3 and the SRDEF file).
XC	A plus (+) sign indicates that the CUI2 for this row has children in this context. If this field is empty, the CUI2 does not have children in this context.

Sample Records

C0002871|S0013742|MSH2002|D000740|1|ANC|1|MeSH|C0220876||||
C0002871|S0013742|MSH2002|D000740|1|ANC|2|Diseases (MeSH Category)|C0012674|C|||
C0002871|S0013742|MSH2002|D000740|1|ANC|3|Hemic and Lymphatic Diseases|C0018981|C15|||
C0002871|S0013742|MSH2002|D000740|1|ANC|4|Hematologic Diseases|C0018939|C15.378|isa||
C0002871|S0013742|MSH2002|D000740|1|CCP||Anemia|C0002871|C15.378.71|isa|+|
C0002871|S0013742|MSH2002|D000740|1|CHD||Anemia, Aplastic|C0002874|C15.378.71.85|isa|+|
C0002871|S0013742|MSH2002|D000740|1|SIB||Blood Protein Disorders|C0005830|C15.378.147|isa|+|
C0002871|S0013742|MSH2002|D000740|1|CHD||Anemia, Hemolytic|C0002878|C15.378.71.141|isa|+|

2.7.1.2.11 Associated Expressions (File = MRATX)

There is one row in this table for each vocabulary expression (i.e., combination of terms from a specific Metathesaurus source vocabulary) identified as having a relationship to a concept in the Metathesaurus. The majority of Metathesaurus entries have no entries in this table.

Col.	Description
CUI	Unique identifier of concept to which the expression is related
SAB	Abbreviation of source of terms in expression. Allowed values are listed in Appendix B, Section B.1)
REL	Relationship of meaning of expression to main concept
ATX	Associated expression

Sample Records

C0001207|MSH2002|S|<Acromegaly> AND <Gigantism>|
C0001296|LCH90|U|<Insurance>/<Statistics>|
C0001355|MSH2002|S|<Kidney Failure, Acute> AND <Kidney Papillary Necrosis>|

2.7.1.2.12 Locators (File = MRLO)

Selected information sources in which the Metathesaurus concept was detected.

Notes:

NLM is reviewing these references to external sources of information and would appreciate feedback from users about the utility or suggested changes to this file.

HSTAT was removed from this file in 2001, since current information is better obtained and searched at:

http://hstat.nlm.nih.gov

There is one row in this table for each Metathesaurus concept identified as appearing in each of a selected set of a machine-readable information sources. If the same concept is identified as appearing in more than one of these information sources (e.g., MEDLINE, DXPLAIN, QMR), it will have multiple rows in this table.

These columns are described in the appendix:

Col.	Description
CUI	Unique identifier of concept
ISN	Name of information source or database in which concept appears
FR	Frequency count of number of occurrences of concept in the information source (optional)
UN	Meaning of frequency (optional)
SUI	Unique identifier of string if name used in information source appears in MRCON (optional)
SNA	Actual name that occurs in the information source if not otherwise present in the Metathesaurus (optional)
SOUI	Unique identifier of record in which the concept appears in source (optional)

Sample Records

```
C0002871|DXP94|||S0352787|||
C0002871|MBD02|2361|*CITATIONS|S0013742|||
C0002871|MED02|898|*CITATIONS|S0013742|||
C0002871|PDQ2001|||S0414880|||
C0002871|QMR96|||S0013787|||
C0002872|QMR96||||Anemias||
```

The Metathesaurus does not contain all concepts appearing in some of these information sources, and some Metathesaurus concepts present in these sources may not be flagged as such if the concept is represented in the information source by a term or concept name not present in the Metathesaurus. In some cases the source name is followed by: the frequency of occurrence in that source, the precise meaning of the frequency data, (e.g., whether it refers to a number of citations, etc.). Information in this element can be used to determine which of a limited repertoire of sources contain detected information closely related to the concept. System developers are encouraged to add data for local information sources to this element to support local applications. Data for additional nationally available sources may be included in subsequent versions of the Metathesaurus.

For MEDLINE segments only, if the frequency subelement is blank, the term did not appear as a principal point in any MEDLINE citations for that segment.

2.7.1.2.13 Concept Name Ranking (File=MRRANK)

There is exactly one row for each concept name type from each Metathesaurus source vocabulary (each SAB-TTY combination). The RANK and SUPRES values in the distributed file are those used in Metathesaurus production. Users are free to change these values to suit their needs and preferences, then change the naming precedence and

suppressibility (TS in MRCON) by using MetamorphoSys to create a customized Metathesaurus.

Col.	Description
RANK	Numeric order of precedence, higher value wins
SAB	Abbreviation for source vocabulary
TTY	Abbreviation for concept name type in source vocabulary
SUPRES	Flag indicating that this SAB and TTY will create a TS=s MRCON entry; see TS

Sample Records

0210|AIR93|SY|N|
0209|ULT93|PT|N|
0208|CPT2002|PT|N|

2.7.1.2.14 Word Index (File = MRXW.BAQ, MRXW.DAN, MRXW.DUT, MRXW.ENG, MRXW.FIN, MRXW.FRE, MRXW.GER, MRXW.HEB, MRXW.HUN, MRXW.ITA, MRXW.NOR, MRXW.POR, MRXW.RUS, MRXW.SPA, MRXW.SWE)

There is one row in these tables for each word found in each unique Metathesaurus string (ignoring upper-lower case). All Metathesaurus entries have entries in the word index. The entries are sorted in ASCII order.

Col.	Description
LAT	Abbreviation of language of the string in which the word appears
WD	Word in lowercase
CUI	Concept identifier
LUI	Term identifier
SUI	String identifier

Sample Records from MRXW.ENG

ENG|anaemia|C0002871|L0280031|S0352688|
ENG|anemia|C0002871|L0002871|S0013742|
ENG|anemias|C0002871|L0002871|S0013787|
ENG|blood|C0002871|L0376533|S0500659|
ENG|cells|C0002871|L0376533|S0500659|

Sample Records from MRXW.FRE

FRE|ANEMIE|C0002871|L0162748|S0227229|

2.7.1.2.15 Normalized Word Index (File=MRXNW.ENG)

There is one row in this table for each normalized word found in each unique English-language Metathesaurus string. All English-language Metathesaurus entries have entries in the normalized word index. There are no normalized string indexes for other languages in this edition of the Metathesaurus.

Col.	Description
LAT	Abbreviation of language of the string in which the word appears (always ENG in this edition of the Metathesaurus)
NWD	Normalized word in lowercase (described in Section 2.6.2.1)
CUI	Concept identifier
LUI	Term identifier
SUI	String identifier

Sample Records

ENG|anaemia|C0002871|L0280031|S0352688|
ENG|anemia|C0002871|L0002871|S0013742|
ENG|anemia|C0002871|L0002871|S0013787|
ENG|blood|C0002871|L0376533|S0500659|
ENG|cell|C0002871|L0376533|S0500659|

2.7.1.2.16 Normalized String Index (File=MRXNS.ENG)

There is one row in this table for each normalized string found in each unique English-language Metathesaurus string (ignoring upper-lower case). All English-language Metathesaurus entries have entries in the normalized string index. There are no normalized word indexes for other languages in this edition of the Metathesaurus.

Col.	Description
LAT	Abbreviation of language of the string (always ENG in this edition of the Metathesaurus)
NSTR	Normalized string in lowercase (described in Section 2.6.3.1)
CUI	Concept identifier
LUI	Term identifier
SUI	String identifier

Sample Records

ENG|anaemia|C0002871|L0280031|S0352688|
ENG|anaemia unspecified|C0002871|L0696700|S0803315|
ENG|anemia|C0002871|L0002871|S0013787|

2.7.1.2.17 Ambiguous Term Identifiers (File = AMBIG.LUI)

There is exactly one row in this table for each Lexical Unique Identifier (LUI) that is linked to multiple Concept Unique Identifiers (CUIs); i.e., it identifies those lexical variant classes which have multiple meanings in the Metathesaurus.

In the Metathesaurus, the LUI links all strings within the English language that are identified as lexical variants of each other by the luinorm program found in the UMLS SPECIALIST Lexicon and Tools (see sections 4 and 4.8 in this manual). LUIs are assigned irrespective of the meaning of each string. This table may be useful to system developers who wish to make use of the lexical programs in their applications.

Col.	Description
LUI	Lexical Unique Identifier
CUIs	List of Concept Unique Identifiers to which the LUI is linked, separated by commas, e.g., C#######,C#######

2.7.1.2.18 Ambiguous String Identifiers (File=AMBIG.SUI)

There is exactly one row in this file for each string identifier (SUI) that is linked to multiple concept identifiers (CUI). This file is now in the META directory (use to be in CHANGE directory). In the Metathesaurus, there is only one SUI for each unique string within each language, even if the string has multiple meanings. This table is only of interest to system developers who make use of the SUI in their applications or in local data files.

Col.	Description
SUI	String Unique Identifier
CUIs	List of Concept Unique Identifiers to which the SUI is linked, separated by commas, e.g., C#######,C#######

2.7.1.2.19 Metathesaurus Change Files

2.7.1.2.19.1 Introduction

There are six files or relations that identify key differences between entries in the previous and the current edition of the Metathesaurus. They are in the subdirectory called CHANGE in META directory and are also accessible on the UMLS Knowledge Source Server. Developers can use these special files to determine whether there have been changes that affect their applications.

The usefulness of individual files will depend on how data from the Metathesaurus have been linked or incorporated in a particular application.

2.7.1.2.19.2 General Description of the Files

Each relation or named table of data has a fixed number of columns and variable number of rows. A column is a sequence of all the values in a given data element. A row contains the values for two or more data elements for one entry. The values for the different data elements in the row are separated by vertical bars (|). Each row ends with a vertical bar and line termination.

2.7.1.2.19.3 Deleted Concepts (File=DELETED.CUI)

There is exactly one row in this table for each reviewed concept that was present in the previous Metathesaurus and is not present in the 2002AA Metathesaurus.

Cols.
CUI Concept Unique Identifier in the previous Metathesaurus
STR Preferred name of this concept in the previous Metathesaurus

2.7.1.2.19.4 Merged Concepts (File=MERGED.CUI)

There is exactly one row in this table for each released concept in the previous Metathesaurus (CUI1) that was merged into another released concept from the previous Metathesaurus (CUI2). When this merge occurs, the first CUI (CUI1) was retired; this table shows the CUI (CUI2) for the merged concept in this Metathesaurus.

Entries in this file represent concepts pairs that were considered to have different meanings in the previous edition, but which are now identified as synonyms

Cols.
CUI1 Concept Unique Identifier in the previous Metathesaurus
CUI2 Concept Unique Identifier in this Metathesaurus in format C#######

2.7.1.2.19.5 Deleted Terms (File=DELETED.LUI)

There is exactly one row in this table for each Lexical Unique Identifier (LUI) that appeared in the previous Metathesaurus, but does not appear in this Metathesaurus.

Metathesaurus Lexical Unique Identifiers (LUIs) are assigned by the luinorm program, part of LVG program in the UMLS SPECIALIST Lexicon and Tools; see Sections 4 and 4.8 in this manual.

These entries represent the cases where LUIs identified by the previous release's luinorm program, when used to identify lexical variants in the previous Metathesaurus, are no longer found with this release's luinorm on this release's Metathesaurus. This does not necessarily imply the deletion of a string or a concept from the Metathesaurus.

Cols.
LUI Lexical Unique Identifier in the previous Metathesaurus
STR Preferred Name of Term in the previous Metathesaurus

2.7.1.2.19.6 Merged Terms (File=MERGED.LUI)

There is exactly one row in this file for each case in which strings had different Lexical Unique Identifiers (LUIs) in the previous Metathesaurus yet share the same LUI in this Metathesaurus; a LUI present in the previous Metathesaurus is therefore absent from this Metathesaurus.

Metathesaurus Lexical Unique Identifiers (LUIs) are assigned by the luinorm program, part of the LVG program in the UMLS SPECIALIST Lexicon and Tools; see Sections 4 and 4.8 in this manual.

These entries represent the cases where separate lexical variants as identified by the previous release's luinorm program version are a single lexical variant as identified by this release's luinorm.

Cols.
LUI Lexical Unique Identifier in the previous Metathesaurus but not present in this Metathesaurus
LUI Lexical Unique Identifier into which it was merged in this Metathesaurus

2.7.1.2.19.7 Deleted Strings (File=DELETED.SUI)

There is exactly one row in this file for each string in each language that was present in a entry in the previous Metathesaurus and does not appear in this Metathesaurus.

Note that this does not necessarily imply the deletion of a term (LUI) or a concept (CUI) from the Metathesaurus. A string deleted in one language may still appear in the Metathesaurus in another language.

Cols.
SUI String Unique Identifier in previous Metathesaurus that is not present in this Metathesaurus
LAT Three character abbreviation of language of string that has been deleted.

STR Preferred name of term in previous Metathesaurus that is not present in this Metathesaurus.

2.7.1.2.19.8 Retired CUI Mapping (File=MRCUI)

There are one or more rows in this file for each Concept Unique Identifier (CUI) that existed in any prior release but is not present in the current release. The file includes mappings to current CUIs as synonymous or to one or more related current CUI where possible. If a synonymous mapping cannot be found, other relationships between the CUIs can be created. These relationships can be Broader (RB), Narrower (RN) or Other Related (RO). Some CUIs may be mapped to more than one other CUI using these relationships.

CUIs may be retired when (1) two released concepts are found to be synonyms and so are merged, retiring one CUI; (2) when the concept no longer appears in any source vocabulary and is not 'rescued' by NLM; or (3) where the concept is an acknowledged error in a source vocabulary or determined to be a Metathesaurus production error.

See the META/CHANGE files, especially MERGED.CUI and DELETED.CUI, for the changes from the last release only, without mappings.

Col.	Description
CUI1	Retired CUI - was present in some prior release, but is currently missing
VER	The last release version in which CUI1 was a valid CUI
CREL	The relationship CUI2 has to CUI1, if present, or DEL if CUI2 is not present. Valid values currently are SY,DEL, RO, RN, RB
CUI2	The current CUI that CUI1 most closely maps to.

Sample Records:
C0079138|2001|DEL||
C0079138|2001|RO|C0037440|
C0079151|1993|DEL||
C0079158|1997|SY|C0009081|
C0079167|1997|SY|C0010042|

2.7.1.2.19.9 Large Scale Vocabulary Test (LSVT) Identifier Mappings (File=XUI.MAP)

(NOTE: For 2001, the new version of XUI.MAP will only be available for LSVT researchers to ftp from the UMLS Knowledge Source Server. It will not be released on the CDs.)

There is exactly one row in this file for each string in SNOMED International, the Read Thesaurus, and LOINC that was part of the planned additions file used in the NLM/AHCPR Large Scale Vocabulary Test and is now part of the 2002 Metathesaurus. Those who did not participate in the LSVT can ignore this file.

Cols.
XUI in LSVT planned additions file
LUI in LSVT planned additions file
SUI in LSVT planned additions file
CUI in 2002 Metathesaurus
LUI in 2002 Metathesaurus
SUI in 2002 Metathesaurus

2.8 MetamorphoSys

2.8.1 Introduction

If you would like to see a tutorial on Customizing the Metathesaurus including how to use MetamorphoSys, please see:

http://umlsinfo.nlm.nih.gov/education.html

MetamorphoSys is a tool to customize and create subsets of the UMLS Metathesaurus for two purposes:

1. to exclude vocabularies as required for your use, under the UMLS License Agreement

2. to alter the Metathesaurus to better meet users' needs by excluding entire vocabularies, altering naming precedence or adding suppressibility to any source-termgroup or CUI|SUI combination.

To select vocabularies to exclude, check the License Agreement and Appendix A.1 of the documentation for a description of each vocabulary.

Note, for example, that CPT2002 (the AMA's Physicians' Current Procedural Terminology, CPT4) is also a part of HCPT02 (the Health Care Financing Administration Common Procedure Coding System, HCPCS). Both vocabularies must be removed to exclude all sources of CPT information.

MetamorphoSys includes a self-contained Java program for configuration and Java subsetting. It has been tested to run on the following operating systems:

Sun Solaris 2.5.1
Windows XP
Windows NT 4.0, Windows 2000

Windows 98, Windows ME
Red Hat Linux 7.1 (Intel x86)

We have successfully tested the Linux version of MetamorphoSys under Red Hat Linux 7.1. The Linux version comes with the 1.3 version of the Java2 JRE. The Java2 v1.3.0 for Linux requires a kernel with GLIBC 2.1.2 or greater.

This program takes from less than one hour to many hours to run on the systems tested (Sun Ultra 2 and PCs ranging from 200 to 866 MHz). Obviously, additional memory, large swap space, faster cpus, and faster disks help.

2.8.2 Space Requirements

The MetamorphoSys system requires that the full UMLS distribution be present on disk. Once you have uncompressed it, simply cd to the UMLS2002AA/META/METAMSYS directory and start the MetamorphoSys script as detailed below.

Space requirements for the full UMLS is approximately 3.25 GB. The full Metathesaurus occupies 3.00 GB. The subsetted Metathesaurus could (in the worst case of the smallest source being removed) take roughly another 3.00 GB, so total space requirements are 6.0 GB assuming no overhead.

We recommend a minimum of 8 GB free disk space to use MetamorphoSys. On NT systems the disk must be formatted as a NT file system (NTFS), since the limit on the partition size for FAT file systems is 2GB. For the same reason, Windows98 disks must be FAT32.

We recommend 256 MB of physical memory to run MetamorphoSys. Less will be likely to cause virtual memory swapping to disk and very slow performance; requirements depend to some extent on the number of concepts and strings in the subset, with smaller subsets requiring less memory.

2.8.3 How to Run

The MetamorphoSys system includes the Java Runtime Environment for each platform supported. A windowing system (not provided) is required to run the Java interface.

To start MetamorphoSys, simply go to the UMLS2002AA/META/METAMSYS/ directory and execute the MetamorphoSys program (MetamorphoSys.sh in the UNIX and Linux environments, MetamorphoSys.bat in Windows; do not execute from other directories). The system will start a Java graphical user interface (window) that interactively prompts the user for information.

The interface is composed of four "tabs":

The "Files/folders" tab allows the user to specify the source folder (directory) for the UMLS installation and the target folder where the subsetted files will be placed. The user can type the directory paths into the text boxes if the defaults are not desired. Alternatively, the user can "Browse" the file system structure by clicking on the appropriate button next to each listed folder. Additionally, the user can change the current configuration file by either typing into the "Current Configuration File" text box or by clicking on the "Browse" button next to that text box.

The "Sources" tab can be brought to the front by clicking on it. This tab allows users to specify which sources (vocabularies) to exclude. Note that all sources with a Restriction Level greater than zero have been preselected for exclusion. Use mouse click to add another source to the selection, or to de-select a source. Two options in the Advanced Options window, described below, provide the capability to select multiple related sources for exclusion with a single click.

The "Precedence" tab allows the user to select the (single) highest precedence source. The default highest precedence source is the Metathesaurus itself (MTH).

The "Term Status" tab allows the user to select which additional source and term type combinations to make suppressible. Users may then delete or treat suppressible terms differently in their applications.

The "Sources", "Precedence", and "Term Status" tables discussed above can all be sorted by the data in any particular table column when the column's header is selected. Additionally, all tables can be reverse sorted when a mouse click is performed on the column's header.

The Options menu allows the user to reset the default selections for the "Sources", "Precedence", or "Term Status" tables. The default selections are those listed in the mmsys.prop.default file in the config folder. The mmsys.prop.sav file contains the properties used in the last run of MetamorphoSys.

The Options menu also contains an Advanced Options window which is described below in the Advanced Users section.

The user can choose to exit the program by the "Exit" option in the File menu.

Once the desired choices have been entered, the process of subsetting can be started by selecting the "Begin MetamorphoSys..." option from the File menu. A window pops up, advising the user that the interface will disappear and the subsetting process will run for several hours.

A progress monitor charts the subsetting progress. First the progress of the initializing CUI list is shown. Then the progress in subsetting the MR Files is shown and that is followed by the progress of the index file subsetting. Pressing cancel on the progress monitor window will cause the application to abort.

When the subsetting process is complete, simple progress messages and any errors encountered are displayed to a the screen and written to a log file called "mmsys.log" in the directory containing the subsetted files.

After the program completes, the 'subsetted' Metathesaurus files should be present in the chosen directory (UMLS2002AA/METASUBSET by default).

**Note that the data in MRRANK, MRCUI, MRFILES and MRCOLS are not recomputed for the subsetted Metathesaurus, they are simply passed through "as is". Also, the STT field in the subsetted MRCON will not exactly match the ones in the original as these are computed without the full LVG machinery.

2.8.4 Advanced Users

Selecting the "Advanced Options" item from the "Options" menu opens a configuration dialog which contains the following user capabilities.

1. Remove Suppressible Data - If the "Remove Suppressible Data" check-box is selected, all data in which the term status ts='s' MRCON will be removed from the result set. For example, the following row would be among those removed. C0000731|ENG|s|L0658950|PF|S0835542|Change in abd size/distension|0|

   The corresponding rows will be removed from other files containing the same CUI,SUI. If this operation causes all rows for a CUI to be removed in MRCON, that entire CUI will be excluded from the result set for the other files. The default for this flag is false.

2. Enforce Family Selection - If the "Enforce Family Selection" check-box is selected, the user will be prompted to select other sources that share the same "Source Family" as the source that the user clicked.

   For example, if RCD is selected, the user will be provided with a list containing RCDAE, RCDSA, and RCDSY and will be given the opportunity to select those sources to be excluded as well because they share RCD as their Source Family. This works for deselection as well. To continue the example, if RCD is deselected, the user is given the opportunity to deselect RCDAE, RCDSA, and RCDSY as well. The default for this flag is true.

3. Suppress Preferred Terms - If the "Suppress Preferred Terms" check-box and the "Remove Suppressible Data" check-boxes are selected, suppressible term status ts='P' rows in MRCON will be removed from the result set. A ts='P' row is suppressible if either its Source Abbreviation/Term Type is in list of selected rows from the Term Status

table or if its CUI,SUI appears in the "suppressed_terms" property in the configuration file.

For example, if MSH2002|PM was selected on the "Suppressed Termgroups" table, then the following row would be among those removed from the result set.
C0000052|ENG|P|L0000052|PF|S0007584|1,4-alpha-Glucan Branching Enzyme|0| The default for this flag is false.

4. Enforce Dependent Source Selection - If the "Enforce Dependent Source Selection" check-box is selected and the user selects a source that appears in the "Dependent Source Associations" table in the "Advanced Options" window, the user will then be offered the option to additionally select any sources listed as dependent sources in that table. As with "Enforce Family Selection" this functionality exists for deselection of sources as well. The default for this flag is true.

The user can choose to add Source/Dependent Source relationships to the "Dependent Source Associations" table by clicking on the "Add" button. The user may clear the whole table by clicking on the "Clear" button. The user may remove a specific line or lines from the table by selecting those lines and pressing the "Delete" button. The user may also sort the table either by clicking on the Source or "Dependent Source" table header. The user may do a reverse sort of the table by pressing while clicking on a table header. The user may exit the "Advanced Options" dialog by clicking on the "Done" button at the bottom of the window.

5. Auto Enforce Family and Dependent Source Selection - If the "Auto Enforce Family and Dependent Source Selection" check box is selected, the user is not prompted when the selected row shares a Source Family or has a Dependent Source. The system selects the Dependent Source rows or the rows with the same Source Family automatically. The default for this flag is false.

The "Precedence" tab allows the user to select a source to outrank all other sources. This has the effect of ranking all of that source's term types above the term types from the other sources, while preserving the ranking order shown in the tab for the other sources.

Advanced users may want to entirely re-rank the term types shown rather than just select the highest ranking source. Selecting the "Edit Precedence" item from the "Options" menu opens a dialog window that allows this kind of re-ordering. A list of source abbreviations and term types is shown and the user can cut and paste the rows into the desired order. Once the "Done" button is selected, this order is shown in the "Precedence" tab and is preserved for the actual subsetting operation. The "Cancel" button cancels the operation.

In order to ensure that the correct versions of sources are removed during subsetting, it is required that the MetamorphoSys distribution release match the release of the data. The release version of MetamorphoSys can be found on the title bar of the application. This release version should match the version in the release.dat file in the directory above the installation directory where the MR files are located. Otherwise, it should match the name of the directory above the installation directory.

2.8.5 Getting Help

Check the information available at:

http://umlsinfo.nlm.nih.gov

and especially:

http://umlsinfo.nlm.nih.gov/mmsys.html

While the MetamorphoSys is not a supported NLM product, NLM does maintain a listserv (electronic mailing list service) called umls-users where requests for help may be sent.

To subscribe to the listserv, send a message to

listserv@nlm.nih.gov

which includes the following line:

subscribe umls-users

To post a message to the umls-users listserv AFTER subscribing, send email to:

umls-users@lhc.nlm.nih.gov

We are also developing web resources for UMLS users; start your searches at:

http://www.nlm.nih.gov/research/umls/

2.8.6 Acknowledgments

Solaris and Windows Java Runtime Environment: http://www.javasoft.com

Linux Java Runtime Environment: http://www.blackdown.org

SECTION 3
UMLS SEMANTIC NETWORK

3.0 Introduction

The UMLS Semantic Network provides a consistent categorization of all concepts represented in the UMLS Metathesaurus and the important relationships between them. This section of the documentation provides an overview of the Semantic Network, as well as information about the Network's semantic types and semantic relationships. The files that contain this information are described, and sample records give the reader an indication of the structure and content of each of the files.

3.1 Overview

The purpose of the Semantic Network is to provide a consistent categorization of all concepts represented in the UMLS Metathesaurus and to provide 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 2002AA release of the Semantic Network contains 134 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 relationships 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.

The Metathesaurus consists of terms from its controlled source vocabularies. The meaning of each term is defined by its source, explicitly by definition or annotation; by context (its place in a hierarchy); by synonyms and other stated relationships between terms; and by its usage in description, classification, or indexing. Each Metathesaurus concept is assigned at least one semantic type. In all cases, the most specific semantic type available in the hierarchy is assigned to the concept. For example, the concept "Macaca" receives the semantic type " Mammal" because there is not a more specific type "Primate" available in the Network. The level of granularity varies across the Network. This has important implications for interpreting the meaning (i.e., semantic type) that has been assigned to a Metathesaurus concept. For example, a sub-tree under the node "Physical Object" is "Manufactured Object". It has only two child nodes, "Medical Device" and "Research Device". It is clear that there are manufactured objects other than medical devices and research devices. Rather than proliferate the number of semantic types to encompass multiple additional subcategories for these objects, concepts

that are neither medical devices nor research devices are simply assigned the more general semantic type "Manufactured Object".

Figure 1 illustrates a portion of the Network. The semantic type "Biologic Function" has two children, "Physiologic Function" and "Pathologic Function", and each of these in turn has several children and grandchildren. Each child in the hierarchy is linked to its parent by the "isa" link.

A Portion of the UMLS Semantic Network

"Biologic Function" Hierarchy

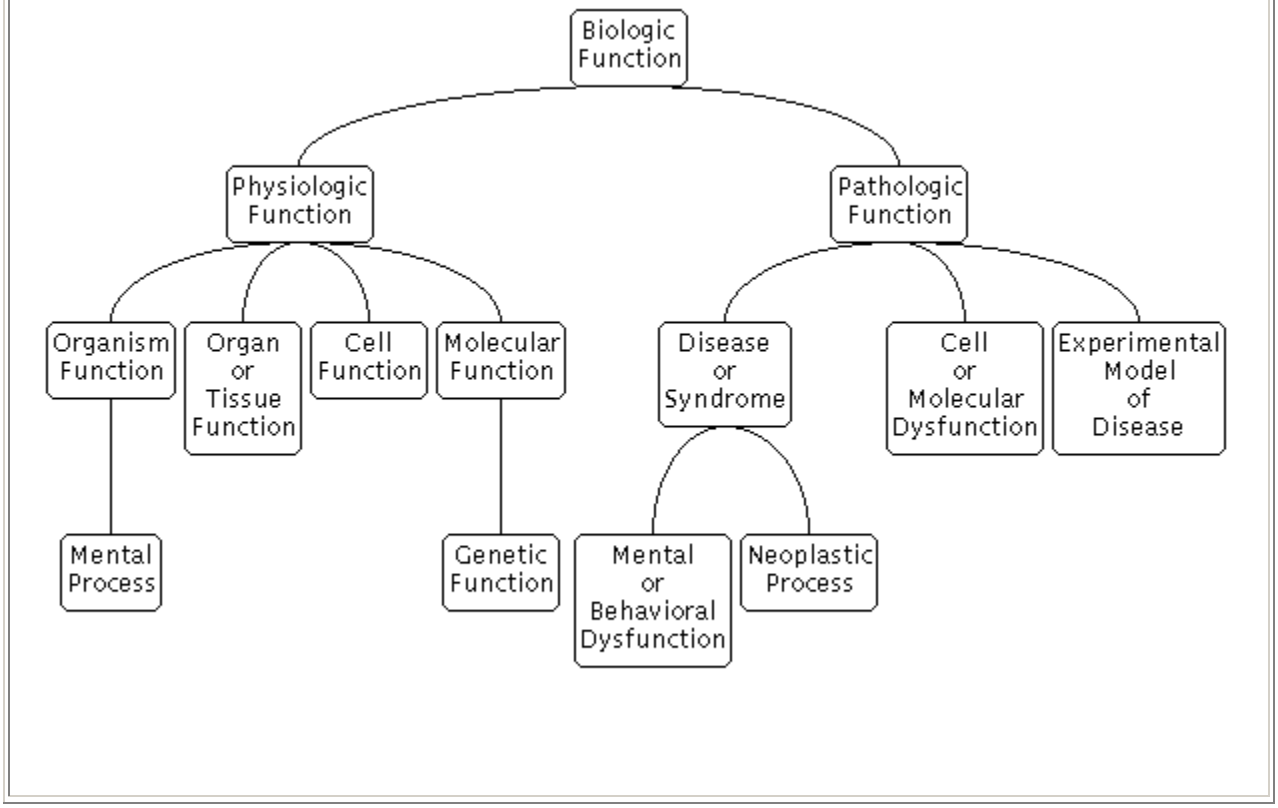


Figure 1

The primary link in the Network is the "isa" link. This establishes the hierarchy of types within the Network and is used for deciding on the most specific semantic type available for assignment to a Metathesaurus concept. In addition, a set of non-hierarchical relations between the types has been identified. These are grouped into five major categories, which are themselves relations: "physically related to", "spatially related to", "temporally related to", "functionally related to", and "conceptually related to".

Figure 2 illustrates a portion of the hierarchy for Network relationships. The "affects" relationship, one of several functional relationships, has six children, including "manages", "treats", and "prevents".

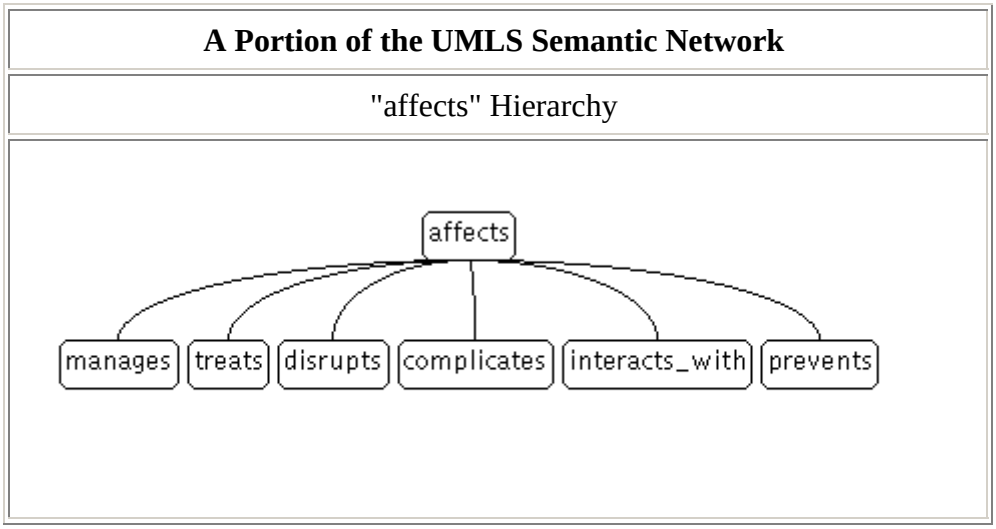


Figure 2

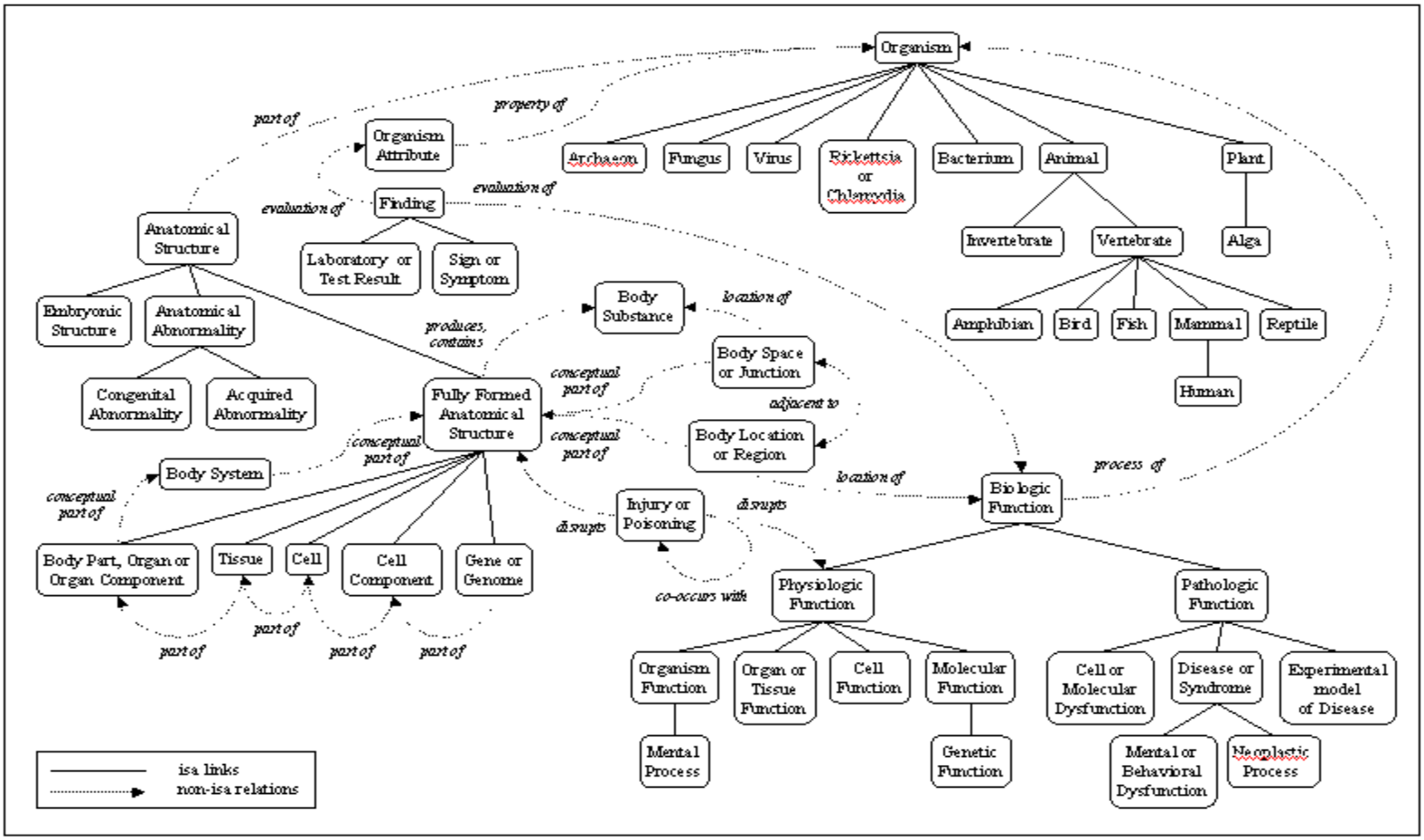
The relations are stated between high level semantic types in the Network whenever possible and are generally inherited via the "isa" link by all the children of those types. Thus, for example, the relation "process of" is stated to hold between the semantic types "Biologic Function" and "Organism". Therefore, it also holds between "Organ or Tissue Function" (which is a "Physiologic Function", which is, in turn, a "Biologic Function") and "Animal" (which is an "Organism"). The relations are stated between semantic types and do not necessarily apply to all instances of concepts that have been assigned to those semantic types. That is, the relation may or may not hold between any particular pair of concepts. So, though the relation "evaluation of" holds between the semantic types "Sign" and "Organism Attribute", a particular sign or a particular attribute may not be linked by this relation. Thus, signs such as "overweight" and "fever" are evaluations of the organism attributes "body weight" and "body temperature", respectively. However, "overweight" is not an evaluation of "body temperature", and "fever" is not an evaluation of "body weight".

In some cases there will be a conflict between the placement of types in the Network and the link to be inherited. If this is so, the inheritance of the link is said to be blocked. For example, by inheritance, the type "Mental Process" would be "process of" "Plant". Since plants are not sentient beings, this link is explicitly blocked. In other cases the nature of the relation is such that it should not be inherited by the children of the types that it links. In that case, the relation is defined for the two semantic types it explicitly links, but blocked for all the children of those types. For example, "conceptual part of" links "Body System" and "Fully Formed Anatomical Structure", but it should not link "Body System" to all the children of "Fully Formed Anatomical Structure", such as "Cell" or "Tissue".

Several portions of the MeSH hierarchy have been labeled with child to parent semantic relationships. All of the anatomy, diseases, and psychiatry and psychology sections have been labeled, as well as a portion of the biological sciences section. The links that are expressed between MeSH terms are, with a few exceptions, reflected in the Semantic Network. That is, if two MeSH terms are linked by a certain relation, then that link is expressed in the Network as a link between the semantic types that have been assigned to those MeSH terms. For example, "Amniotic Fluid", which is a "Body Substance", is a child of "Embryo", which is an "Embryonic Structure". The labeled relationship between "Amniotic Fluid" and its parent "Embryo" is "surrounds". This is allowable, since the relation "Body Substance surrounds Embryonic Structure" is represented in the Network.

Figure 3 shows a portion of the Semantic Network, illustrating the relations, either hierarchical or associative, that exist between semantic types.

A Portion of the UMLS



Semantic Network

relations

Figure 3

The UMLS Semantic Network is provided in two formats: a relational table format and a unit record format.

3.2 Semantic Network ASCII Relational Format

There are two basic tables, two ancillary tables, and two bookkeeping tables included in this format. The two basic tables contain exactly the same information as the unit record file, but the information is presented somewhat differently. One table contains definitional information about the semantic types and relations; the other contains information about the structure of the Network. Each semantic type and each relation has been assigned a four character unique identifier (UI). These are of the form "T001", "T002", etc. The ancillary tables are expansions of the table that contains the Network structure. They give the fully inherited set of links represented in the Network. The first table is expressed as triples of UI's. The second is expressed as triples of names. The two bookkeeping tables describe the relational files and their fields. Fields in all tables are separated by a "|". All tables are listed and described below:

Table	Description
SRDEF	Basic information about the Semantic Types and Relations.
SRSTR	Structure of the Network.
SRSTRE1	Fully inherited set of Relations (UI's).
SRSTRE2	Fully inherited set of Relations (names).
SRFIL	Description of each table.
SRFLD	Description of each field and the table(s) in which it is found.

Specific Descriptions of each Table:

Table: SRDEF

RT:	Record Type (STY = Semantic Type or RL = Relation).
UI:	Unique Identifier of the Semantic Type or Relation.
STY/RL:	Name of the Semantic Type or Relation.
STN/RTN:	Tree Number of the Semantic Type or Relation.
DEF:	Definition of the Semantic Type or Relation.
EX:	Examples of Metathesaurus concepts with this Semantic Type (STY records only).
UN:	Usage note for Semantic Type assignment (STY records only).
NH:	The Semantic Type and its descendants allow the non-human flag (STY records only).
ABR:	Abbreviation of the Relation Name (RL records only).
RIN:	Inverse of the Relation (RL records only).

Table: SRSTR

STY/RL:	Argument 1 (Name of a Semantic Type or Relation).
RL:	Relation ("isa" or the name of a non-hierarchical Relation).
STY/RL:	Argument 2 (Name of a Semantic Type or Relation); if this field is blank this means that the Semantic Type or Relation is one of the top nodes of the Network.
LS:	Link Status (D = Defined for the Arguments and its children; B = Blocked; DNI = Defined but Not Inherited by the children of the Arguments). N.B.: The relations expressed in this table are binary relations and the arguments are ordered pairs. The relations are stated only for the top-most node of the "isa" hierarchy of the Semantic Types to which they may apply.

Table: SRSTRE1 or SRSTRE2

UI/STY:	Argument 1 (UI or name of a Semantic Type).
UI/RL:	Relation (UI or name of a nonhierarchical Relation).
UI/STY:	Argument 2 (UI or name of a Semantic Type). N.B.: The relations expressed in this table are binary relations and the arguments are ordered pairs. All relations have been fully inherited in this table.

Table: SRFIL

FIL:	File Name.
DES:	Description of the file.
FMT:	Format of the file (fields in a comma-separated list).
CLS:	Number of columns in the file.
RWS:	Number of rows in the file.
BTS:	Number of bytes in the file.

Table: SRFLD

COL:	Field name.
DES:	Description of the field.
REF:	Cross-reference to the documentation.
FIL:	File name(s) in which the field is found.

Sample Relational Records

```
::::::::::::::
SRDEF
::::::::::::::
STY|T020|Acquired Abnormality|A1.2.2.2|An abnormal structure,
or one that is abnormal in size or location, found in or deriving
from a previously normal structure. Acquired abnormalities are
distinguished from diseases even though they may result in pathological
functioning (e.g., "hernias incarcerate").|Abscess of
prostate; Hemorrhoids; Hernia, Femoral; Varicose Veins|||||

STY|T052|Activity|B1|An operation or series of operations that
an organism or machine carries out or participates in.|Social
Planning; Expeditions; Information Distribution; Return Migration|Few
concepts will be assigned to this broad type. Wherever possible,
one of the more specific types from this hierarchy will be chosen.
 For concepts assigned to this type, the focus of interest is
on the activity. When the focus of interest is the individual
or group that is carrying out the activity, then a type from the
'Behavior' hierarchy will be chosen. In general, concepts will
not receive a type from both the 'Activity' and the 'Behavior'
hierarchies.||||

STY|T100|Age Group|A2.9.4|An individual or individuals classified
according to their age.|Adult; Infant, Premature; Adolescents;
Aged, 80 and over|||||

STY|T003|Alga|A1.1.1.1|A chiefly aquatic plant that contains
chlorophyll,
but does not form embryos during development and lacks vascular
tissue.|Chlorella; Laminaria; Seaweed|||||

RL|T173|adjacent_to|R2.2|Close to, near or abutting another physical
unit with no other structure of the same kind intervening. This
includes adjoins, abuts, is contiguous to, is juxtaposed, and
is close to.||||AD|adjacent_to|

RL|T151|affects|R3.1|Produces a direct effect on. Implied
is the altering or influencing of an existing condition, state,
situation, or entity. This includes has a role in, alters, influences,
predisposes, catalyzes, stimulates, regulates, depresses, impedes,
enhances, contributes to, leads to, and modifies.||||AF|affected_by|

::::::::::::::
SRSTR
::::::::::::::
Acquired Abnormality|co-occurs_with|Injury or Poisoning|D|
Acquired Abnormality|isa|Anatomical Abnormality|D|
Acquired Abnormality|result_of|Behavior|D|
Activity|isa|Event|D|
Age Group|isa|Group|D|
Alga|isa|Plant|D|
```

::::::::::::::
SRSTRE1
::::::::::::::
T020|T186|T190|
T020|T186|T017|
T020|T186|T072|
T052|T186|T051|
T052|T165|T090|
T052|T165|T091|
T100|T186|T096|
T100|T186|T077|
T100|T186|T071|
T003|T186|T002|
T003|T186|T001|
T003|T186|T072|

::::::::::::::
SRSTRE2
::::::::::::::
Acquired Abnormality|isa|Anatomical Abnormality|
Acquired Abnormality|isa|Anatomical Structure|
Acquired Abnormality|isa|Physical Object|
Acquired Abnormality|isa|Entity|
Acquired Abnormality|affects|Alga|
Acquired Abnormality|affects|Amphibian|
Acquired Abnormality|affects|Animal|
Acquired Abnormality|affects|Bacterium|
Acquired Abnormality|affects|Bird|
Acquired Abnormality|affects|Cell Function|
Acquired Abnormality|affects|Fish|
Acquired Abnormality|affects|Fungus|
Acquired Abnormality|affects|Genetic Function|
Acquired Abnormality|affects|Human|
Acquired Abnormality|affects|Invertebrate|
Acquired Abnormality|affects|Mammal|
Acquired Abnormality|affects|Mental Process|
Acquired Abnormality|affects|Molecular Function|
Acquired Abnormality|affects|Organ or Tissue Function|
Acquired Abnormality|affects|Organism Function|
Acquired Abnormality|affects|Organism|
Acquired Abnormality|affects|Physiologic Function|
Acquired Abnormality|affects|Plant|
Acquired Abnormality|affects|Reptile|
Acquired Abnormality|affects|Rickettsia or Chlamydia|
Acquired Abnormality|affects|Vertebrate|
Acquired Abnormality|affects|Virus|
Activity|isa|Event|
Age Group|isa|Group|
Age Group|isa|Conceptual Entity|
Age Group|isa|Entity|
Alga|isa|Plant|
Alga|isa|Organism|
Alga|isa|Physical Object|
Alga|isa|Entity|

3.3 Semantic Network ASCII Unit Record Format

The file "SU" contains individual records for both semantic types and relations.

Each record begins with a unique identifier field (UI) which contains the four character UI. These are of the form "T001", "T002", etc. Each field in a record begins on a new line and may continue over several lines. Some fields are optional.

Semantic Type records contain the following fields:

Field	Description
UI:	Unique Identifier of the Semantic Type.
STY:	Name of the Semantic Type.
STN:	Tree Number of the Semantic Type.
DEF:	Definition of the Semantic Type.
EX:	Examples of Metathesaurus concepts with this Semantic Type (optional field).
UN:	Usage note for Semantic Type assignment (optional field).
NH:	Semantic Type and its descendants allow the non-human flag (optional field).
HL:	Hierarchical links of the Semantic Type to its parent({isa})and its children ({inverse_isa}). If there are no hierarchical links, then the value <none> is assigned.

Relation records contain the following fields:

Field	Description
UI:	Unique Identifier of the Relation.
RL:	Name of the Relation.
ABR:	Abbreviation of the Relation.
RIN:	Name of the inverse of the Relation.
RTN:	Tree Number of the Relation.
DEF:	Definition of the Relation.
INH:	"N" if the relation is not inherited (optional field).
HL:	Hierarchical links of the Relation to its parent ({isa}) and its children ({inverse_isa}). If there are no hierarchical links, then the value <none> is assigned.
STL:	Semantic Types linked by this Relation. N.B.: These are binary relations and the arguments are ordered pairs. The relations are stated only for the top-most node of the "isa"

hierarchy of the Semantic Types to which they may apply. This field does not appear in the "isa" relation record since its values can be computed from the "HL" field. If there are no semantic types linked by this Relation, then the value <none> is assigned.

STLB: Semantic Types linked by this Relation are blocked (optional field).

Sample Unit Records

```
::::::::::::::
SU
::::::::::::::
UI:  T020
STY: Acquired Abnormality
STN: A1.2.2.2
DEF: An abnormal structure, or one that is abnormal in size or
location,
     found in or deriving from a previously normal structure. Acquired
     abnormalities are distinguished from diseases even though they
may
     result in pathological functioning (e.g., "hernias incarcerate").
EX:  Abscess of prostate; Hemorrhoids; Hernia, Femoral; Varicose Veins
HL:  {isa} Anatomical Abnormality
UI:  T052
STY: Activity
STN: B1
DEF: An operation or series of operations that an organism or machine
     carries out or participates in.
EX:  Social Planning; Expeditions; Information Distribution; Return
Migration
UN:  Few concepts will be assigned to this broad type. Wherever
possible, one of the
     more specific types from this hierarchy will be chosen. For
concepts assigned
     to this type, the focus of interest is on the activity. When the
focus of
     interest is the individual or group that is carrying out the
activity, then a
     type from the 'Behavior' hierarchy will be chosen. In general,
concepts will
     not receive a type from both the 'Activity' and the 'Behavior'
hierarchies.
HL:  {isa} Event;
     {inverse_isa} Behavior;
     {inverse_isa} Daily or Recreational Activity;
     {inverse_isa} Occupational Activity;
     {inverse_isa} Machine Activity
UI:  T100
STY: Age Group
STN: A2.9.4
DEF: An individual or individuals classified according to their age.
EX: Adult;
     Infant, Premature; Adolescents; Aged, 80 and over
HL:  {isa} Group
```

```
UI:  T003
STY: Alga
STN: A1.1.1.1
DEF: A chiefly aquatic plant that contains chlorophyll, but does not
form embryos
     during development and lacks vascular tissue.
EX:  Chlorella; Laminaria; Seaweed
HL: {isa} Plant
UI:  T173
RL:  adjacent_to
ABR: AD
RIN: adjacent_to
RTN: R2.2
DEF: Close to, near or abutting another physical unit with no other
structure of
     the same kind intervening. This includes adjoins, abuts, is
contiguous to,
     is juxtaposed, and is close to.
HL:  {isa} spatially_related_to
STL:
     [Body Location or Region|Body Location or Region];
     [Body Location or Region|Body Part, Organ, or Organ Component];
     [Body Location or Region|Body Space or Junction];
     [Body Part, Organ, or Organ Component|Body Part, Organ, or Organ
Component];
     [Body Part, Organ, or Organ Component|Body Space or Junction];
     [Body Part, Organ, or Organ Component|Cell];
     [Body Part, Organ, or Organ Component|Tissue];
     [Body Space or Junction|Body Space or Junction];
     [Cell Component|Body Space or Junction];
     [Cell Component|Cell Component];
     [Cell|Cell];
     [Tissue|Body Space or Junction];
     [Tissue|Tissue]
UI:  T151
RL:  affects
ABR: AF
RIN: affected_by
RTN: R3.1
DEF: Produces a direct effect on. Implied here is the altering or
influencing of
     an existing condition, state, situation, or entity. This includes
has a role
     in, alters, influences, predisposes, catalyzes, stimulates,
regulates,
     depresses, impedes, enhances, contributes to, leads to, and
modifies.
HL:  {isa} functionally_related_to;
     {inverse_isa} manages;
     {inverse_isa} treats;
     {inverse_isa} disrupts;
     {inverse_isa} complicates;
     {inverse_isa} interacts_with;
     {inverse_isa} prevents
STL:
        [Natural Phenomenon or Process|Natural Phenomenon or Process];
        [Anatomical Abnormality|Physiologic Function];
```

```
        [Biologic Function|Organism];
        [Anatomical Abnormality|Organism];
        [Health Care Activity|Biologic Function];
        [Diagnostic Procedure|Patient or Disabled Group];
        [Therapeutic or Preventive Procedure|Patient or Disabled
Group];
        [Chemical|Natural Phenomenon or Process];
        [Gene or Genome|Physiologic Function];
        [Cell Component|Physiologic Function];
        [Physiologic Function|Organism Attribute];
        [Food|Biologic Function];
        [Behavior|Behavior];
        [Behavior|Mental Process];
        [Mental Process|Behavior];
        [Mental or Behavioral Dysfunction|Behavior];
        [Research Activity|Mental Process];
        [Regulation or Law|Group];
        [Regulation or Law|Organization]
```

3.4 Hierarchies for Semantic Types and Relations In the Semantic Network

Current Semantic Types:

Entity
- Physical Object
  - Organism
    - Plant
      - Alga
    - Fungus
    - Virus
    - Rickettsia or Chlamydia
    - Bacterium
    - Archaeon
    - Animal
      - Invertebrate
      - Vertebrate
        - Amphibian
        - Bird
        - Fish
        - Reptile
        - Mammal
          - Human
  - Anatomical Structure
    - Embryonic Structure
    - Anatomical Abnormality
      - Congenital Abnormality
      - Acquired Abnormality
    - Fully Formed Anatomical Structure
      - Body Part, Organ, or Organ Component
      - Tissue
      - Cell
      - Cell Component
      - Gene or Genome
  - Manufactured Object
    - Medical Device
    - Research Device
    - Clinical Drug

[Entity] (continued)
- [Physical Object] (continued)
  - Substance
    - Chemical
      - Chemical Viewed Functionally
        - Pharmacologic Substance
          - Antibiotic
        - Biomedical or Dental Material
        - Biologically Active Substance
          - Neuroreactive Substance or Biogenic Amine
          - Hormone
          - Enzyme
          - Vitamin
          - Immunologic Factor
          - Receptor
        - Indicator, Reagent, or Diagnostic Aid
        - Hazardous or Poisonous Substance
      - Chemical Viewed Structurally
        - Organic Chemical
          - Nucleic Acid, Nucleoside, or Nucleotide
          - Organophosphorus Compound
          - Amino Acid, Peptide, or Protein
          - Carbohydrate
          - Lipid
            - Steroid
            - Eicosanoid
        - Inorganic Chemical
        - Element, Ion, or Isotope
    - Body Substance
    - Food

[Entity] (continued)
- Conceptual Enity
  - Idea or Concept
    - Temporal Concept
    - Qualitative Concept
    - Quantitative Concept
    - Functional Concept
      - Body System
    - Spatial Concept
      - Body Space or Junction
      - Body Location or Region
      - Molecular Sequence
        - Nucleotide Sequence
        - Amino Acid Sequence
        - Carbohydrate Sequence
      - Geographic Area
  - Finding
    - Laboratory or Test Result
    - Sign or Symptom
  - Organism Attribute
    - Clinical Attribute
  - Intellectual Product
    - Classification
    - Regulation or Law
  - Language
  - Occupation or Discipline
    - Biomedical Occupation or Discipline
  - Organization
    - Health Care Related Organization
    - Professional Society
    - Self-help or Relief Organization
  - Group Attribute
  - Group
    - Professional or Occupational Group
    - Population Group
    - Family Group
    - Age Group
    - Patient or Disabled Group

Event
- Activity
  - Behavior
    - Social Behavior
    - Individual Behavior
  - Daily or Recreational Activity
  - Occupational Activity
    - Health Care Activity
      - Laboratory Procedure
      - Diagnostic Procedure
      - Therapeutic or Preventive Procedure
    - Research Activity
      - Molecular Biology Research Technique
    - Governmental or Regulatory Activity
    - Educational Activity
  - Machine Activity
- Phenomenon or Process
  - Human-caused Phenomenon or Process
    - Environmental Effect of Humans
  - Natural Phenomenon or Process
    - Biologic Function
      - Physiologic Function
        - Organism Function
          - Mental Process
        - Organ or Tissue Function
        - Cell Function
        - Molecular Function
          - Genetic Function
      - Pathologic Function
        - Disease or Syndrome
          - Mental or Behavioral Dysfunction
          - Neoplastic Process
        - Cell or Molecular Dysfunction
        - Experimental Model of Disease
  - Injury or Poisoning

Current relations in the Semantic Network:

- isa
- associated with
  - physically related to
    - part of
    - consists of
    - contains
    - connected to
    - interconnects
    - branch of
    - tributary of
    - ingredient of
  - spatially related to
    - location of
    - adjacent to
    - surrounds
    - traverses
  - functionally related to
    - affects
      - manages
      - treats
      - disrupts
      - complicates
      - interacts with
      - prevents
    - brings about
      - produces
      - causes

- [associated with] (continued)
  - [functionally related to] (continued)
    - performs
      - carries out
      - exhibits
      - practices
    - occurs in
      - process of
    - uses
    - manifestation of
    - indicates
    - result of
  - temporally related to
    - co-occurs with
    - precedes
  - conceptually related to
    - evaluation of
    - degree of
    - analyzes
      - assesses effect of
    - measurement of
    - measures
    - diagnoses
    - property of
    - derivative of
    - developmental form of
    - method of
    - conceptual part of
    - issue in

SECTION 4

THE SPECIALIST LEXICON AND LEXICAL PROGRAMS

4.0 Introduction

The SPECIALIST lexicon has been developed to provide the lexical information needed for the SPECIALIST Natural Language Processing System (NLP). It 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 NLP 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, and "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 this sort of variation.

For an overview of the SPECIALIST lexicon, lexical variant programs, and lexical databases, see "Lexical Methods for Managing Variation in Biomedical Terminologies", A.T. McCray, S. Srinivasan, A.C. Browne, in the Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care, 1994, 235-239.

4.1 General Description

The lexicon consists of a set of lexical entries with one entry for each spelling or set of spelling variants in a particular part of speech. Lexical items may be "multi-word" terms made up of other words if the multi-word term is determined to be a lexical item by its presence as a term in general English or medical dictionaries, or in medical thesauri such as MeSH. Expansions of generally used acronyms and abbreviations are also allowed as multi-word terms.

The unit lexical record is a frame structure consisting of slots and fillers. Each lexical record has a base= slot whose filler indicates the base form, and optionally a set of spelling_variants= slots to indicate spelling variants. An "entry=" slot records the unique identifier (EUI) of the record. EUI numbers are seven digit numbers preceded by an "E". Each record has a cat= slot indicating part of speech. The lexical record is delimited by braces ({...}).

The unit lexical records for "anaesthetic" given below illustrate some of the features of the SPECIALIST lexical record:

```
{base=anesthetic
spelling_variant=anaesthetic
entry=E0330018
        cat=noun
        variants=reg
        variants=uncount
}
{base=anesthetic
spelling_variant=anaesthetic
entry=E0330019
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
}
```

The base form "anesthetic" and its spelling variant "anaesthetic" appear in two lexical records, one an adjective entry, the other a noun entry. The variants= slot contains a code indicating the inflectional morphology of the entry; the filler reg in the noun entry indicates that the noun "anesthetic" is a count noun which undergoes regular English plural formation ("anaesthetics"); inv in the variants= slot of the adjective entry indicates that the adjective "anesthetic" does not form a comparative or superlative. The position= slot indicates that the adjective "anaesthetic" is attributive and appears after color adjectives in the normal adjective order. "pred" in the position slot of the adjective entry indicates that this adjective can appear in predicate position.

Lexical entries are not divided into senses. Therefore, an entry represents a spelling-category pairing regardless of semantics. The noun "act" has two senses both which show a capitalized and lower case spelling; an act of a play and an act of law. Since both senses share the same spellings and syntactic category, they are represented by a single lexical entry in the current lexicon. The unit record for "Act" is shown below.

```
{base=Act
spelling_variant=act
entry=E0000154
        cat=noun
        variants=reg
}
```

When different senses have different syntactic behavior, codes for each behavior are recorded in a single entry. For example, "beer" has two senses: the alcoholic beverage and the amount of a standard container of that beverage.

A. Patients who drank beer recovered more slowly than patients who drank wine.
B. Fifty-six patients reported drinking more than five beers a day.

The first sense illustrated in A. above is a mass (uncount) noun. The second sense illustrated in B. is a regular (count) noun. In cases like this the appropriate codes for both senses are included in the entry.

```
{base=beer
entry=E0012226
        cat=noun
        variants=uncount
        variants=reg
}
```

Two codes will also appear in cases where the lexical item is both count and uncount without a sense distinction. "Abdominal delivery" denotes the same procedure whether it appears as an uncount noun as in C. or a count noun as in D.

C. Abdominal delivery is the procedure of choice in this situation.
D. Abdominal deliveries are more common these days.

The unit lexical record for "abdominal delivery" includes both codes.

```
{base=abdominal delivery
entry=E0006453
        cat=noun
        variants=uncount
        variants=reg
signature=stavri
}
```

Other syntactic codes such as complement codes for verbs, adjectives and nouns are similarly grouped without regard to sense.

4.2 The Scope of the Lexicon

Words are selected for lexical coding from a variety of sources. Approximately 20,000 words from the UMLS Test Collection of MEDLINE abstracts together with words which appear both in the UMLS Metathesaurus and Dorland's Illustrated Medical Dictionary form the core of the words entered. In addition, an effort has been made to include words from the general English vocabulary. The 10,000 most frequent words listed in The American Heritage Word Frequency Book and the list of 2,000 words used in definitions in Longman's Dictionary of Contemporary English have also been coded. Since the majority of the words selected for coding are nouns, an effort has been made to include verbs and adjectives by identifying verbs in current MEDLINE citation records, by using the Computer Usable Oxford Advanced Learner's Dictionary, and by identifying potential adjectives from Dorland's Illustrated Medical Dictionary using heuristics developed by McCray and Srinivasan (1990).

A variety of reference sources are used in coding lexical records. Coding is based on actual usage in the UMLS Test Collection and MEDLINE, dictionaries of general English, primarily learner's dictionaries which record the kind of syntactic information needed for NLP, and medical dictionaries. Longman's Dictionary of Contemporary English, Dorland's Illustrated Medical Dictionary, Collins COBUILD Dictionary, The Oxford Advanced Learner's Dictionary, and Webster's Medical Desk Dictionary were used.

The SPECIALIST lexicon also exists in an ASCII relational format generated from the unit records. The full SPECIALIST lexicon technical report entitled "The SPECIALIST Lexicon", found in the file techrpt.pdf, fully describes the unit record format. The remainder of the present section describes the ASCII relational form of the lexicon. Section 4.3 describes the Data elements that make up the relational tables and Section 4.4 describes the tables.

4.3 Lexicon Data Elements

Each of the elements below are represented as fields (columns) in the ASCII relational format.

4.3.1 String Properties

These data elements refer to properties of the strings generated by the entries.

4.3.1.1 STR - String

A Lexical entry generates a variety of forms (strings) including all the inflectional forms (the citation form, as well) of each spelling variant. Case, punctuation and spaces are considered significant.

4.3.1.2 AGR - Agreement/Inflection Code

This element encodes agreement and inflection information.

Agreement between nouns and verbs and between determiners and nouns involves person and number. Person and Number are indicated by the following codes.

Code	Person	Number
second	Second	Singular & Plural
third	Third	Singular & Plural
fst_sing	First	Singular
fst_plur	First	Plural
thr_sing	Third	Singular
thr_plur	Third	Plural

For Nouns, the agreement/inflection code indicates countability, person and number. Person and number are indicated by the person/number codes given above which are parenthesized after the countability code. Nouns can be either count or uncount.

For Pronouns, the agreement/inflection indicates person and number using the codes given above.

For verbs, including auxiliaries and modals, the agreement/inflection code indicates tense, person and number. Persons and numbers are indicated by the same person/number codes given above. These codes are parenthesized after the tense. No person number codes are given for non-finite tenses. "pres(thr_sing)" indicates third person singular present tense and "pres(fst_sing,fst_plur,thr_plur,second)" indicates present tense for all persons and numbers other than third singular. Negative forms of auxiliaries (didn't) and modals (can't) have "negative" after a colon at the end of the agreement/inflection code.

Code	Tense
past	Past Tense
pres	Present Tense
past_part	Past Participle
pres_part	Present Participle
infinitive	Infinitive

Determiners agree with nouns in terms of countability and number. The agreement/inflection codes for determiners are "free", "plur", "sing" and "uncount". "free" indicates that the determiner places no restrictions on it's noun. Determiners marked "plur" allow plural nouns, those marked "sing" allow singular nouns and those marked "uncount" allow uncount nouns.

4.3.1.3 CAS - Case

See Section 4.3.1 of "The SPECIALIST Lexicon".

Pronouns in English may be in one of two cases, subjective (nominative) or objective (accusative). This field contains "subj", "obj" or both separated by a comma to indicate the case of the pronoun.

4.3.1.4 GND - Gender

This field indicates the gender of pronouns.

Pronouns may be marked pers or neut to indicate whether they refer to people or non-people respectively. Pronouns marked pers may be masculine (masc) or feminine (fem) referring to male or female people respectively. See Section 14.2 of "The SPECIALIST Lexicon". There are four codes possible in this field:

Code	Gender
pers	person
neut	neuter
pers(masc)	person masculine
pers(fem)	person feminine

Notice that pers as used here does not correspond to the traditional term "personal pronoun". For example "it" and "they" are traditionally called personal pronouns since they both participate in the person/number paradigm. A pronoun like "none" is not traditionally called a personal pronoun.

4.3.2 Entry Properties

4.3.2.1 EUI - Unique Identifier Number for Lexical Entries

The EUI identifies a lexical entry. Information about a set of spelling variants in a particular part of speech is represented as an entry in the unit record. A particular string may be assigned several EUI numbers as it may occur in several parts of speech.

4.3.2.2 CIT - Citation Form

This field records the citation form of strings in the agreement/inflection table (lragr). The citation form is the singular for nouns, infinitive for verb and positive for adjectives and adverbs. The base form and the spelling variants if any are the citation forms of each of their respective inflections. This form is sometimes referred to as the un-inflected form.

4.3.2.3 BAS - Base Form

This field records the base form of a lexical entry. The base form is the citation form of one of a set of spelling variants chosen to represent the whole set. It might be thought of as the name of a lexical entry. The base form is the filler of the base= slot.

4.3.2.4 SCA - Syntactic Category

The syntactic category (part of speech) of the lexical entry. This field may be filled by one of the following. See Section 3 of "The SPECIALIST Lexicon".

Code	Category
noun	nouns
adj	adjectives
adv	adverbs
pron	pronouns
verb	verbs
det	determiners
prep	prepositions
conj	conjunctions
aux	auxiliaries
modal	modals
compl	complementizers

4.3.2.5 PER - Periphrastic

The code "periph" in this field indicates that an adjective or adverb is periphrastic. An adjective is periphrastic if it can form its comparative with "more" and its superlative with "most". See Section 4.3.5 of "The SPECIALIST Lexicon" for discussion.

4.3.2.6 COM - Complements

These are complement codes. See Sections 5.1, 5.2, 5.4 and 5.5 in "The SPECIALIST Lexicon" for a description of SPECIALIST complement codes.

4.3.2.7 TYP - Inflectional Type

The inflectional type(s) of an entry indicate the ways in which it's forms may be inflected, or in the case of determiners the inflection of the heads they may determine. These codes are used to generate the variant strings (STR) found in other tables.

For nouns the following types may appear:

Code	Pluralization Pattern	See "The SPECIALIST Lexicon" Section
reg	regular	4.5.2
glreg	Greco-Latin regular	4.5.3

metareg	metalinguistic regular	4.5.4
irreg( )	irregular	4.5.5
sing	fixed singular	4.5.6
plur	fixed plural	4.5.7
inv	invariant	4.5.8
group(irreg( ))	group irregular	4.5.9
group(reg)	group regular	4.5.9
uncount	uncountable	4.5.10
groupuncount	group uncount	4.5.11

For verbs the following types may appear:

Code	Inflection Type	See "The SPECIALIST Lexicon" Section
reg	regular	4.1.1
regd	regular doubling	4.1.2
irreg( )	irregular	4.1.3

For pronouns the following types may appear:

Code	Inflection Type
fst_plur	first person plural
fst_sing	first person singular
sec_plur	second person plural
sec_sing	second person singular
second	second person
third	third person
thr_plur	third person plural
thr_sing	third person singular

See Section 14.1 of "The SPECIALIST Lexicon".

For adjectives and adverbs the following types can appear:

Code	Inflectional Type	See "The SPECIALIST Lexicon" Section
reg	regular	4.3.1 and 4.4.1
regd	regular doubling	4.3.2
inv	invariant	4.3.4 and 4.4.3
inv;periph	periphrastic	4.3.5 and 4.4.4
irreg( )	irregular	4.3.3 and 4.4.2

For determiners the infection type indicates the inflection of the noun heads they may determine. The following types may appear:

Code	Inflectional Type	See "The SPECIALIST Lexicon" Section
sing	singular	4.7.1
plur	plural	4.7.2
uncount	uncount	4.7.3
singuncount	singular uncount	4.7.4
pluruncount	plural uncount	4.7.5
free	free	4.7.6

4.3.2.8 POS - Possession

English pronouns may be possessive, or possessive nominal. The codes poss, possnom or both (comma separated) may appear in this field.

See Section 14.3.2 of "The SPECIALIST Lexicon".

4.3.2.9 QNT - Quantification

This field indicates the quantification properties inherent in certain pronouns. The four codes possible in this field are:

Code	Properties
univ	universal quantification
indef(nonassert)	non-assertive indefinite

indef(neg)	negative indefinite
indef(assert)	assertive indefinite

See Section 14.3.4 in "The SPECIALIST Lexicon" for discussion of quantification in pronouns.

4.3.2.10 FEA - Features

This field represents various features of terms in various categories. The possible features are:

Feature	Section of "The SPECIALIST Lexicon"
reflexive	14.3.3
negative	14.3.4
demonstrative	14.3.5
interrogative	12.1
proper	8.
negative	13.1
broad_negative	13.2
stative	10.

4.3.2.11 PSN - Position for Adjectives

Adjectives are marked in the SPECIALIST lexicon with position codes showing whether they are attributive postmodifying or predicative. If attributive, the code indicates where they appear in the pre-nominal sequence of adjectives. An additional attributive code, attribc, is used to indicate adjectives which can take complements in attributive position. One or more of the following codes can appear:

Code	Position	See "The SPECIALIST Lexicon" Section
attrib(1)	attributive (1st position)	9.1.1.1
attrib(2)	attributive (2nd position)	9.1.1.2
attrib(3)	attributive (3rd position)	9.1.1.3

attribc	attributive with complement	9.1.2
post	post modifying	9.2
pred	predicative	9.3

4.3.2.12 MOD - Modification Type for Adverbs

Adverbs are marked in the SPECIALIST lexicon to indicate their modification type. The possible values of this field are:

Code	See "The SPECIALIST Lexicon" Section
intensifier	11.2
particle	11.1
sentence_modifier; TYPE	11.3
verb_modifier; TYPE	11.4

TYPE is one of locative, temporal or manner. See Section 11.5 in "The SPECIALIST Lexicon".

4.3.2.13 GEN - Generic Name for a Trademark

The GEN field represents a generic or public name for the thing referred to by the trademark. The trademark "Alphalin" has the generic term "vitamin A".

4.3.3 Entry Relations

4.3.3.1 ABR - Acronym or Abbreviation

This field indicates whether a term listed in the acronym-abbreviation table (lrabr) is an acronym or abbreviation. It contains either:

"abbreviation_of" or "acronym_of".

4.3.3.2 SPV - Spelling Variant

A base form in the SPECIALIST lexicon may have one or more spelling variants, subject to the same inflectional pattern. This field contains the citation form of a particular

spelling variant. See Section (2) of "The SPECIALIST Lexicon".

4.3.4 Data Description

The data elements describe the relational table files or provide index entries into the lexicon.

4.3.4.1 WRD - Word

Each string is broken into "words" and indexed in lrwd. Words are strings of alpha-numeric characters more than one character long, separated by space or punctuation.

4.3.4.2 DES - Description

A short definition of a file or field. This is free text.

4.3.4.3 FMT - Format

An ordered comma separated list of field names appearing in a file.

4.3.4.4 RWS - Number of Rows

The number of Rows (lines or records) in a file.

4.3.4.5 FIL - File Name(s)

One or more file names denoting the files containing relational tables.

4.3.4.6 BTS - Size in Bytes

The size of a file in bytes (characters).

4.3.4.7 CLS - Number of Columns

The number of columns (fields) in a record (or row) of a table. The same number as the number of lines in the file.

4.3.4.8 COL - Three Letter Field Name

A three letter identifier for a field.

4.3.4.9 REF - Cross Reference to Document

A cross reference to a section of this document.

4.4 Lexicon ASCII Relational Tables

4.4.1 Introduction

In this format the data in each lexical entry is represented in ten different "relations" or "tables" each in a file.

The Lexicon relational format is not fully normalized. By design, there is duplication of data among different relations and within certain relations. Developers will need to make their own decisions about the extent to which this redundancy should be retained, reduced, or increased for their specific applications.

4.4.2 General Description of the Relational Format

As in the Metathesaurus ASCII relational format, each relation or table of data values has by definition a fixed number of columns; the number of rows depends on the content of a particular version of the Lexicon. A column is a sequence of all the values in a given data element or logical sub-element. In general, columns for longer variable length data elements will appear to the right of columns for shorter and/or fixed length data elements. A row contains the values for one or more data elements or logical sub-elements for one Lexicon entry or string. Depending on the nature of the data elements involved, each Lexicon entry or string may have one or more rows in a given file. The values for the different data elements or logical sub-elements represented in the row are separated by vertical bars (|). If an optional element is blank, the vertical bars are still used to maintain the correct positioning of the subsequent elements. Each row is terminated by a vertical bar and a carriage return followed by a line feed. (|<CR><LF>).

4.4.3 Summary of the Contents of Each of the Relational Files

In the following descriptions, the numbers in parentheses beside each element refer to the section of this document that describes the element's contents.

4.4.3.1 - Agreement and Inflection (File = lragr)

Rows of the agreement table have six fields. There is a row in lragr for each inflected form of each spelling variant. This table links those forms to their citation forms and base forms. It provides information about agreement between subjects (nouns and pronouns) and verbs and between determiners and nouns.

EUI	The Entry Unique ID Number (4.3.2.1)
STR	String (4.3.1.1)
SCA	Syntactic Category (4.3.2.4)
AGR	Agreement/Inflection Code (4.3.1.2)

CIT	Citation Form (4.3.2.2)
BAS	Base Form (4.3.2.3)

4.4.3.2 - Inflection Type (File = lrtyp)

The lrtyp table has one or more rows for each lexical entry, indicating the inflectional pattern(s) to which it belongs.

EUI	The Entry Unique ID Number (4.3.2.1)
BAS	The Base Form (4.3.2.3)
SCA	Syntactic Category (4.3.2.4)
TYP	Inflectional Type (4.3.2.7)

4.4.3.3 - Complementation (File = lrcmp)

In lrcmp there is one line for each complement code for each entry.

EUI	The Entry Unique ID Number (4.3.2.1)
BAS	The Base Form (4.3.2.3)
SCA	Syntactic Category (4.3.2.4)
COM	Complement Code. (4.3.2.6)

4.4.3.4 - Pronouns (File = lrprn)

lrprn has one or more rows for each pronoun entry in the lexicon. Each row has nine columns.

EUI	The Entry Unique ID Number (4.3.2.1)
BAS	The Base Form (4.3.2.3)
AGR	Agreement/Inflection Code (4.3.1.2)

See Section 14.1 in "The SPECIALIST Lexicon".

The agreement/inflection field in lrprn indicates person and number for anaphoric reference, AGR in lragr indicates person for agreement. These differ in the case of possessive nominal pronouns. The possessive nominal "mine" is "third" for purposes of

subject verb agreement and "fst_sing" in it's anaphoric reference.

GND	Gender (4.3.1.6)
CAS	Case (4.3.1.3)
POS	Possession (4.3.2.8)
QNT	Quantification (4.3.2.9)
FEA	Other Features (for pronouns) (4.3.2.10)

4.4.3.5 Modifiers (file = lrmod)

The modifier table includes position information for adjectives and modification type information for adverbs, and a variety of features.

EUI	The Entry Unique ID Number (4.3.2.1)
BAS	The Base Form (4.3.2.3)
SCA	Syntactic Category (4.3.2.4)

All the entries represented in this table have the category "adj" or "adv" indicating adjectives or adverbs respectively.

PSN/MOD

The fourth field of lrmod may be one of the following depending on whether the term is an adjective or adverb.

PSN	Position (4.3.2.11) - for adjectives
MOD	Modification Types (4.3.2.12) - for adverbs
FEA	Features (4.3.2.10)

4.4.3.6 - Properties (file = lrprp)

lrprp indicates properties of terms in various categories.

EUI	The Entry Unique ID Number (4.3.2.1)
BAS	The Base Form (4.3.2.3)
SCA	Syntactic Category (4.3.2.4)
STR	String (4.3.1.1)

STR is only indicated in lrprp when a feature applies to a single string out of those generated by the entry, as in the negative contractions.

FEA Features (4.3.2.10)

4.4.3.7 - Abbreviations and Acronyms (file = lrabr)

This file links acronyms and abbreviations to their expansions.

EUI The Entry Unique ID Number (4.3.2.1)

This field contains the EUI of the acronym or abbreviation.

BAS The Base Form (4.3.2.3)

This field contains the Base form of the acronym or abbreviation.

ABR Acronym or Abbreviation (4.3.3.1)
BAS The Base Form (4.3.2.3)

This field contains the Base form of the expansion of the acronym or abbreviation.

EUI The Entry Unique ID Number (4.3.2.1)

This field contains the EUI of the expansion of the abbreviation or acronym.

4.4.3.8 - Spelling Variants (lrspl)

EUI The Entry Unique ID Number (4.3.2.1)
SPV Spelling Variant (4.3.3.2)
BAS The Base Form (4.3.2.3)

4.4.3.9 - Nominalizations (file =lrnom)

This field contains the EUI of the nominalization.

BAS The Base Form (4.3.2.3)

This field contains the base form of the nominalization.

SCA Syntactic Category (4.3.2.4)

This field contains the category of the nominalization (noun).

EUI The Entry Unique ID Number (4.3.2.1)

This field contains the EUI of a verb or adjective of which the noun is a nominalization.

BAS The Base Form (4.3.2.3)

This field contains the base form of the verb or adjective of which the noun is a nominalization.

SCA Syntactic Category (4.3.2.4)

This field contains the syntactic category (adj or verb) of the adjective or verb.

4.4.3.10 - Trademarks (file = lrtrm)

EUI The Entry Unique ID Number (4.3.2.1)

BAS Base (4.3.2.3)

GEN Generic Term (4.3.2.13)

The appearance of a form in the lrtrm table indicates that it is a trademark. It may or may not have a generic term associated with it.

4.4.3.11 - Files (file = lrfil)

The lrfil table describes each file in the ASCII relational form of the lexicon.

FIL File Name(s) (4.3.4.5)

DES Description (4.3.4.2)

FMT Format (4.3.4.3)

CLS Number of Columns (4.3.4.7)

RWS Number of Rows (4.3.4.4)

BTS Size in Bytes (4.3.4.6)

4.4.3.12 - Word Index. (file = lrwrd)

WRD	Word (4.3.4.1)
EUI	The Entry Unique ID Number (4.3.2.1)

4.4.3.13 - Fields (file = lrfld)

COL	Three Letter Field Name (4.3.4.8)
DES	Description (4.3.4.2)
REF	Cross Reference to Document (4.3.4.9)
FIL	File Name(s) (4.3.4.5)

4.5 The SPECIALIST Lexicon Unit Record

The unit lexical record is a frame structure consisting of slots and fillers. Each lexical record has a base= slot whose filler indicates the base form, and optionally a set of spelling_variants= slots to indicate spelling variants. Lexical entries are delimited by entry= slots filled by the EUI number of the entry. EUI numbers are seven digit numbers preceded by an "E". Each entry has a cat= slot indicating part of speech. The lexical record is delimited by braces ({...}).

The unit lexical records for "anaesthetic" given below illustrate some of the features of a SPECIALIST unit lexical record:

```
{base=anesthetic
spelling_variant=anaesthetic
entry=E0330018
        cat=noun
        variants=reg
        variants=uncount
}
{base=anesthetic
spelling_variant=anaesthetic
entry=E0330019
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
}
```

The base form "anesthetic" and its spelling variant "anaesthetic" appear in two lexical records containing a noun and a verb entry. The variants= slot contains a code indicating the inflectional morphology of the entry; the filler reg in the noun entry indicates that the noun "anaesthetic" is a count noun which undergoes regular English plural formation

("anaesthetics"); inv in the variants= slot of the adjective entry indicates that the adjective "anesthetic" does not form a comparative or superlative. The position= slot indicates that the adjective "anaesthetic" is attributive and appears after color adjectives in the normal adjective order.

The SPECIALIST Technical report "The SPECIALIST Lexicon" gives a full description of the lexicon in unit format.

4.6 Lexical Databases

4.6.0 Introduction

The lexical databases contain lexical information that we have found to be useful for Natural Language Processing. They are not finished products but are under continuous development.

4.6.1 Semantically Related Terms SM.DB

This database (SM.DB) contains pairs of semantically related terms. Each row of the database has the following form.

TERM1|SCA1|TERM2|SCA2

Such a row indicates that TERM1 in syntactic category SCA1 is semantically related to TERM2 in syntactic category SCA2. Both terms are given in base form.

Examples:

```
alar|adj|wing|noun
ocular|adj|eye|noun
auditory area|noun|auditory cortex|noun
vomitive|noun|emetic|noun
vomitive|adj|emetic|adj
iridescent virus|noun|iridovirus|noun
typhloteritis|noun|cecitis|noun
```

4.6.2 Derivationally Related Terms: DM.DB

This database (DM.DB) contains pairs of terms related by derivational morphology. Each row of the database has the same form as sm.db. Both terms are given in base form.

TERM1|SCA1|TERM2|SCA2

Examples:

```
abashment|noun|abash|verb
adenohypophyseal|adj|adenohypophysis|noun
```

```
amenorrheic|adj|amenorrhea|noun
arithmetician|noun|arithmetic|noun
convert|verb|conversible|adj
immobilize|verb|immobility|noun
```

DM.DB is derived from the morphological fact files (dm.fct, etc.) used in LVG.

4.6.3 Spelling Variants: SP.DB

The Spelling Variant database (SP.DB) contains pairs of terms that are spelling variants of each other. The format of each row is the same as the format of dm.db and sm.db. SCA1 and SCA2 are always the same in SP.DB.

TERM1|SCA1|TERM2|SCA2

Examples:

```
accouter|verb|accoutre|verb
accurst|adj|accursed|adj
acidaemic|adj|acidemic|adj
aesthetics|noun|esthetics|noun
dairy farmer|noun|dairy-farmer|noun
```

SP.DB is derived from the SPECIALIST lexicon.

4.6.4 Neo-classical Combining Forms NC.DB

This database (NC.DB) contains morphemes that are used to form neo-classical compounds. Each row of the database has the following form.

MORPHEME|MEANING|TYPE

Morphemes may have optional connecting vowels indicated in parentheses. The types are: prefix, root, and terminal.

Examples:

```
abdomin(o)|abdomen|root
ab|away from|prefix
acou(o)|hearing|root
cardi(o)|heart|root
cele|swelling|terminal
desis|binding|terminal
de|negate|prefix
```

Our analysis of combining forms divides them into roots and terminals, which are distinguished from prefixes and suffixes. A neo-classical compound can consist of any number of roots ending in a terminal or suffix. Prefixes normally must precede roots and

cannot attach directly to terminals. Users interested in suffixation rules and facts should consult the dm.rul and dm.fct files included with LVG.

For further discussion see McCray et. al., 1988, "The Semantic Structure of Neo-Classical Compounds", In the Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care, Washington DC.

4.7 Sample Records

```
::::::::::::::
lragr.sam
::::::::::::::
E0007127|acute|adj|positive;periph|acute|acute|
E0014875|cans|noun|count(thr_plur)|can|can|
E0014875|can|noun|count(thr_sing)|can|can|
E0014876|canned|verb|past_part|can|can|
E0014876|canned|verb|past|can|can|
E0014876|canning|verb|pres_part|can|can|
E0014876|cans|verb|pres(thr_sing)|can|can|
E0014876|can|verb|infinitive|can|can|
E0014876|can|verb|pres(fst_sing,fst_plur,thr_plur,second)|can|can|
E0014877|can't|modal|pres:negative|can|can|
E0014877|cannot|modal|pres:negative|can|can|
E0014877|can|modal|pres|can|can|
E0014877|couldn't|modal|past:negative|can|can|
E0014877|could|modal|past|can|can|
E0014937|canine teeth|noun|count(thr_plur)|canine tooth|canine tooth|
E0014937|canine tooth|noun|count(thr_sing)|canine tooth|canine tooth|
E0017902|colors|noun|count(thr_plur)|color|color|
E0017902|color|noun|count(thr_sing)|color|color|
E0017902|color|noun|uncount(thr_sing)|color|color|
E0017903|colored|verb|past_part|color|color|
E0017903|colored|verb|past|color|color|
E0017903|coloring|verb|pres_part|color|color|
E0017903|colors|verb|pres(thr_sing)|color|color|
E0017903|color|verb|infinitive|color|color|
E0017903|color|verb|pres(fst_sing,fst_plur,thr_plur,second)|color|color
|
E0051632|quickly|adv|positive;periph|quickly|quickly|
E0055585|she|pron|thr_sing|she|she|

::::::::::::::
lrcmp.sam
::::::::::::::
E0014876|can|verb|tran=np|
E0017903|color|verb|cplxtran=np,adj|
E0017903|color|verb|cplxtran=np,np|
E0017903|color|verb|intran;part(in)|
E0017903|color|verb|intran;part(up)|
E0017903|color|verb|intran|
E0017903|color|verb|tran=np;part(in)|
E0017903|color|verb|tran=np|
```

```
::::::::::::::
lrmod.sam
::::::::::::::
E0007127|acute|adj|attrib(1),attrib(3),pred|stative|
E0051632|quickly|adv|verb_modifier;manner||

::::::::::::::
lrnom.sam
::::::::::::::
E0007121|acuity|noun|E0007127|acute|adj|
E0021126|deduction|noun|E0021123|deduce|verb|
E0021126|deduction|noun|E0021124|deduct|verb|
E0061851|transportation|noun|E0061850|transport|verb|

::::::::::::::
lrprn.sam
::::::::::::::
E0030918|he|thr_sing|pers(masc)|subj||||
E0036100|it|thr_sing|neut|subj,obj||||
E0055585|she|thr_sing|pers(fem)|subj||||

::::::::::::::
lrprp.sam
::::::::::::::
E0007127|acute|acute|adj|stative|
E0004825|Parkinson|Parkinson|noun|proper|
E0014877|can|can't|modal|negative|
E0014877|can|can't|modal|negative|
E0014877|can|couldn't|modal|negative|

::::::::::::::
lrspl.sam
::::::::::::::
E0017902|colour|color|
E0017903|colour|color|
E0008769|anesthetic|anaesthetic|
E0008770|anesthetic|anaesthetic|

::::::::::::::
lrtrm.sam
::::::::::::::
E0412633|Actinex|meso-nordihydroguaiaretic acid|
E0415286|Antivert||
E0414928|thioplex|thiotepa|
E0415019|theo-hexanicit||

::::::::::::::
lrtyp.sam
::::::::::::::
E0007127|acute|adj|inv;periph|
E0014875|can|noun|reg|
E0014876|can|verb|regd|
E0014937|canine tooth|noun|irreg(canine teeth)|
E0017902|color|noun|reg|
E0017902|color|noun|uncount|
E0017903|color|verb|reg|
E0051632|quickly|adv|inv;periph|
```

::::::::::::::
lrwd.sam
::::::::::::::
acute|E0001203
acute|E0007127
acute|E0007130
acute|E0007131
acute|E0007132
acute|E0007133
acute|E0007134
acute|E0007135
acute|E0007136
acute|E0007137
acute|E0007138
acute|E0007139
acute|E0007140
acute|E0007141
acute|E0007142
acute|E0007143
acute|E0007144
acute|E0007145
acute|E0007146
acute|E0007147
acute|E0007148
acute|E0007149
acute|E0007150
acute|E0007151
acute|E0007152
acute|E0007153
acute|E0007154
acute|E0007155
acute|E0007156
acute|E0007157
acute|E0007158
acute|E0007159
acute|E0007160
acute|E0007161
acute|E0007162
acute|E0007163
acute|E0007164
acute|E0007165
acute|E0007166
acute|E0007167
acute|E0007168
acute|E0007169
acute|E0007170
acute|E0007171
acute|E0007172
acute|E0007173
acute|E0007174
acute|E0007175
acute|E0007176
acute|E0007177
acute|E0007178
acute|E0007179
acute|E0007180

acute|E0007181
acute|E0007182
acute|E0007183
acute|E0007184
acute|E0007185
acute|E0007186
acute|E0007187
acute|E0007188
acute|E0007189
acute|E0007190
acute|E0007191
acute|E0007192
acute|E0007193
acute|E0007194
acute|E0007195
acute|E0007196
acute|E0007197
acute|E0007198
acute|E0007199
acute|E0007200
acute|E0007201
acute|E0007202
acute|E0007203
acute|E0007204
acute|E0007205
acute|E0007206
acute|E0007207
acute|E0007208
acute|E0007209
acute|E0007210
acute|E0007211
acute|E0007212
acute|E0007213
acute|E0007214
acute|E0007215
acute|E0016430
acute|E0018044
acute|E0019256
acute|E0200089
acute|E0200090
acute|E0203254
acute|E0208423
acute|E0208433
acute|E0208452
acute|E0208475
acute|E0208494
acute|E0210443
acute|E0210574
acute|E0210575
acute|E0210576
acute|E0210642
acute|E0214476
acute|E0216615
acute|E0216616
acute|E0217176
acute|E0217376
acute|E0217551

```
acute|E0217756
acute|E0313307
acute|E0314926
acute|E0319558
acute|E0321232
acute|E0321304
acute|E0322005
acute|E0332592
acute|E0409630
acute|E0418090
acute|E0418484
acute|E0418485
acute|E0418705
acute|E0420121
acute|E0422597
acute|E0422634
acute|E0422824
acute|E0422825
can|E0014875
can|E0014876
can|E0014877
can|E0014875
can|E0014876
can|E0014877
```

4.8 Lexical Programs

The lexical variant generation package consists of three primary programs -- a normalizer, a word index generator, and a lexical variant generator, together with a set of ancillary programs for normalization.

This package is implemented in Java (V1.2). Updates and bug fixes to this versions may be found at the url: http://umlsks.nlm.nih.gov/KSS/LVG.

The distributions come with install programs (for Solaris, Linux, and Window) and a ReadMe.txt file describing how to install and configure the lexical programs and providing a brief description of each program.

The **docs** directory contains user guides, Java API documents, and design documents describing in detail the use of Lexical tools. This document is a general introduction to the programs in the lexical variant generation package for the 2002 version.

The compressed lexical programs are as follows:

lvg2002.tar.gz

> - The official 2002 distribution of LVG. This includes the source code for the programs, the data and tables in a pure Java embedded database (Instant DB) the programs use, full documentation, installation instructions, and jar files of the programs. See the documents contained within this distribution for a more complete description of this product.

lvg2002.basic.tar.gz

> - The simplified 2002 version of LVG. This includes the source code for the programs, the data the programs use, full documentation, installation instructions, and jar files of the programs. See the documents contained within this distribution for a more complete description of this product.

Normalization (norm)

The lexical program **norm** generates the normalized strings that are used in the normalized string index, MRXNS. Thus norm must be used before MRXNS can be searched.

The normalization process involves stripping possessives, replacing punctuation with spaces, removing stop words, lower-casing each word, breaking a string into its constituent words, and sorting the words in alphabetic order. The uninflected forms are generated using the SPECIALIST lexicon if words appear in the lexicon, otherwise they are generated algorithmically. When a form could be an inflection of more than one base form, the new normalization process returns multiple uninflected forms. If a string to be normalized contains multiple ambiguous forms, and the permutation of these ambiguous forms offer more than 10 output forms, the input form lowercased, with punctuation replaced, word order sorted, but not uninflected, is returned. The upper limit of permutation number (10) is configurable by modifying the configuration file. The program **luiNorm** has the behavior of prior year's normalization, and is distributed for those who need it.

Norm reads its standard input and writes to standard output. It expects input lines to be records separated into fields. The field separator is |. The string to be normalized is identified to norm using the **-t** option. **-t** takes a numerical argument which denotes the field in which the input string is to be found. If no **-t** option appears, norm assumes that the input string is in the first field (**-t:1**). There need not be more than one field, so lines consisting only of input strings are properly understood.

Norm output records include all the fields of the input record with an additional field to the right containing the normalized form of the input string.

For example, if the user had a list of terms to be looked up via the normalized string index in a file called **terms**, he or she could use **norm -i:terms -o:terms.nrm** to get the normalized form of each term. If the input file **terms** contained the following:

```
2, 4-Dichlorophenoxyacetic acid
Syndrome, anterior, compartment
Abnormal, weight, gain
Anemia, Refractory, with Excess of Blasts
left atriums
```

the file **term.nrm** would contain:

```
2, 4-Dichlorophenoxyacetic acid|2 4 acid dichlorophenoxyacetic
Syndrome, anterior, compartment|anterior compartment syndrome
Abnormal, weight, gain|abnormal gain weight
Anemia, Refractory, with Excess of Blasts|anemia blast excess
refractory
left atriums|atrium left
left atriums|atrium leave
```

The string in the second field of each line of terms.nrm is now suitable for matching to MRXNS.

Word Index (wordInd)

The lexical program **wordInd** breaks strings into words for use with the word index in MRXW. Users of the word index should use wordInd to break strings into words before searching in the word index. This assures congruence between the words to be looked up and the word index.

Word for this purpose is defined as a token containing only alphanumeric characters with length one or greater. The wordInd program lowercases the output words.

The wordInd program reads its standard input and writes to its standard output. Like norm and lvg, it expects each input line to be a record separated into fields by |. The field containing the input string is identified using the **-t** option. The numerical argument of **-t** denotes the field in which the input string may be found. If no **-t** option is given, the input string is expected to be in the first field (**-t:1**). There need not be more than one field, so lines consisting only of input strings are properly understood.

The wordInd program outputs one line of output for each word found in the input string. Input fields are not repeated in the output unless specified in a **-F** option. Applying wordInd to the input string **Heart Disease, Acute** would result in three output lines:

```
heart
disease
acute
```

The numerical argument of **-F** indicates an input field to be repeated in the output. A numerical argument for **-F** option is required for each input field that is to be repeated. Fields are repeated in the order in which the numerical argument of **-F** options appear. The output words always appear as an additional field to the right of any repeated input fields. For example, applying **wordInd -t:2 -F:2:1** to a record of the form **UI23456|tooth, canine|definition.....**; would result in the following output:

```
tooth, canine|UI23456|tooth
tooth, canine|UI23456|canine
```

The third field of each of those records contains a word extracted from the input term in the first field (**-t:2 ,-F:2**). The **-F:1** option repeats the UI numbers from the first field of input. The fact that **-F:2:1** placed the UI numbers (field 1) after the input string (field 2).

Lexical Variant Generation (lvg)

The lvg program generates lexical variants of input words. It consists of several different flow components that can be combined in various ways to produce lexical variants. The user of lvg chooses combinations of flow components and combines them into a **flow**. (The normalizer program, norm, is essentially the lvg program with a pre-selected flow option: **lvg -f:N**.) The arguments of the **-f** flag are used to specify a flow. Each flow can be thought of as a pipeline with each flow component feeding the next. For example, the flow **-f:i** simply generates inflectional variants and **-f:l:i** generates lowercase inflectional variants. Each of the flow components options is discussed on the documents for lvg.

The lvg program reads from its standard input and writes to its standard output. Input records may be typed in at the keyboard, after typing the command on the command line (**lvg -f:i**) or input lines may be read from a file (**lvg -f:i -i:file**) or piped to lvg from another command (**COMMAND|lvg -f:i**). Output records may be directed to the screen (default), send to a file (**lvg -f:i -i:INFILE -o:OUTFILE**) or piped to another command (**lvg -f:i -i:infile | COMMAND**).

Input

The lvg program is designed to work with one line input records divided into fields. The default field separator is |. The field separator can be changed using the **-s** option. The field in which the input term, whose variants are to be generated, can be specified with the **-t** option. In the absence of a **-t** flag the input term is assumed to be in the first field of the input. So both **dog** and **dog|canine|UI4567** would generate variants of **dog**. With the **-t** flag set to **2**, **dog|canine|UI4567** would generate variants of **canine**. In the case of single field input (**dog**), lvg generates variants from the only field regardless of the setting of **-t**.

The lvg program can read category (part of speech) and inflection information from the input record. The numerical argument to the **-cf** option indicates the field in which category information is located. In the input record, category information needs to be encoded as a number according to the scheme described on the documents for lvg. The numerical argument to the **-if** option indicates the field in which inflection information is located. In the input record, inflection information needs to be encoded as a number according to the scheme described on the documents for Lexical tools.

Output

The lvg program adds five new fields to the input record and outputs a record for each variant generated. For example, if **dog|canine|UI4567** is given to the standard input of **lvg -f:i** the output sent to standard out will be:

```
dog|canine|UI4567|dog|128|1|i|1|
dog|canine|UI4567|dog|128|512|i|1|
dog|canine|UI4567|dogs|128|8|i|1|
dog|canine|UI4567|dog|1024|1|i|1|
dog|canine|UI4567|dog|1024|262144|i|1|
dog|canine|UI4567|dog|1024|1024|i|1|
dog|canine|UI4567|dogs|1024|128|i|1|
dog|canine|UI4567|dogged|1024|64|i|1|
dog|canine|UI4567|dogged|1024|32|i|1|
dog|canine|UI4567|dogging|1024|16|i|1|
```

The first three fields of each record above are identical to the input record, the rest are supplied by lvg. The first additional field is the variant form lvg has generated. The second additional field is the syntactic category of the variant encoded as a number. The third additional field is the inflection of the variant encoded as a number. The fourth additional field indicates the flow that was selected. The fifth field is the number of the flow which generated this variant. Output category (parts of speech) and inflection information are encoded in the same scheme used for input category and inflection information.

For a more detailed technical discussion of lvg, norm, and wordInd see the documents for Lexical tools.

SECTION 5
USING THE UMLS KNOWLEDGE SOURCE SERVER VIA THE INTERNET

5.0 Background

The UMLS Knowledge Source Server is an evolving tool for providing Internet access to the information stored in the UMLS Knowledge Sources. The purpose of the Knowledge Source Server is to make the UMLS data more accessible to users, and in particular to systems developers. The system architecture is based on the client server paradigm wherein remote site users send their requests to a centrally managed server at the U.S. National Library of Medicine. The client programs can run on platforms supporting the TCP/IP communication protocol. Access to the system is provided through a command line interface, through an Application Programming Interface (API), and through the World Wide Web (WWW).

5.1 Querying the Knowledge Source Server

5.1.1 Metathesaurus

The Knowledge Source Server allows the user to request information about particular Metathesaurus concepts, including attributes such as the concept's definition, its semantic types, the concepts that are related to it, etc. It also allows the user to request information about the attributes themselves; for example, by asking for all the concepts that have been assigned to a particular semantic type.

Basic concept information includes the Metathesaurus unique identifier of the concept, the preferred name for the concept, and the names and sources of all terms that comprise that concept. Additional concept information often includes a definition and the source of that definition. Semantic type information is also included. Information about the hierarchical contexts of Metathesaurus concepts is readily available in the system. Related concepts are easily found. If a user were interested in information about a particular term within a concept, then the results could be limited in that way. Co-occurrence data are included for MeSH and AI-RHEUM terminology.

An important perspective on the Metathesaurus is source specific data. It is possible to query the server by limiting the query to a particular vocabulary. The user may wish to see the ancestors or descendants for a term in just a particular vocabulary, or the user may wish to see just the synonyms for a particular term in a particular vocabulary.

Attributes may be queried in the system. Thus, all concepts with a particular semantic type, all terms with a particular syntactic category, and all terms from a particular source vocabulary may be found. Searching for all concepts with a particular semantic type will, for example, give the user a good idea of the coverage of the Metathesaurus in a subject domain.

5.1.2 Semantic Network

The Semantic Network contains information about semantic types and their relationships. The implementation of the network module computes the relationships between semantic types using the inheritance property of the network type hierarchy. Information in the Semantic Network can be queried for semantic types and the relationships between them. Individual queries are specified by providing the known types or relations and leaving out the unknowns. The system then retrieves the corresponding values for the unknowns. For example, if the user wished to know what semantic types are related by a particular relation, then the user would indicate only the relationship name and all the semantic type pairs linked by that relationship would be retrieved. The user might also wish to know if a particular relationship holds between a pair of types.

It is possible to retrieve all the relations between a pair of types. For example, "treats", "prevents", and "complicates" would be listed, among others, as potential relationships between drugs and diseases. It is also possible to retrieve an exhaustive list of all related types in the network. Queries can be made about the definition, unique identifier, tree number, ancestors, parents, children, descendants, and siblings of a semantic type or relation.

5.1.3 SPECIALIST Lexicon

The Knowledge Source Server provides access to lexical records in the SPECIALIST lexicon. The SPECIALIST lexicon is an English language lexicon containing many biomedical terms. The lexicon entry for each word or term records syntactic, morphological, and orthographic information. Lexical entries may be single or multiword terms. Lexical information includes syntactic category, inflectional variation (e.g., singular and plural for nouns, the conjugations of verbs, the positive, comparative, and superlative for adjectives and adverbs), and allowable complementation patterns (i.e., the objects and other arguments that verbs, nouns, and adjectives can take).

5.2 Gaining Access to the UMLS Knowledge Source Server

Access to the UMLS Knowledge Source Server is available to anyone who has signed the UMLS license agreement and received a license number from NLM. The URL for the Knowledge Source Server Web site is http://umlsks.nlm.nih.gov. First time users should establish a login and a password through the online registration at the web site. Any questions or problems should be addressed via email to umlsks@nlm.nih.gov.

SECTION 6

USING THE UMLS CD-ROMS

6.1 Content of the CD-ROMS

The 2002AA edition of the UMLS Knowledge Sources is available only in compressed CD-ROMs, one for Unix (2002AA_TAR) and one for PC (2002AA_ZIP). The Metathesaurus files have outgrown simple allocation to CD-ROMs in plain ASCII distribution. To use the UMLS, you must uncompress the full distribution to a local hard disk.

Each compressed CD-ROM is in ISO 9660 format and contains the full 2002AA UMLS Knowledge Sources. 2002AA_ZIP contains one file (2002AA.ZIP) in ZIP format with PC line termination in ASCII files. 2002AA_TAR contains 2002AA.TGZ in tar, GNU ZIP (gzip) format (.tar.gz), with Unix line termination in the ASCII files. Appropriate Java Runtime Environments (jre) for MetamorphoSys are included on each CD-ROM.

Use the 2002AA_ZIP CD-ROM for Windows (version 3.1 and up), Windows NT and 2000 (v4.0 and up), and OS/2. PKZIP or WINZIP programs may be used to extract the data; they may be obtained using these URLs: http://www.pkware.com or http://www.winzip.com. To extract the files to your disk, assign the target directory in which you wish to create UMLS2002AA (see below) and unzip. NOTE that your file system must support large file sizes, e.g. NTFS or FAT32.

Use the 2002AA_TAR CD-ROM for operating systems that support UNIX line termination (all flavors of UNIX and Linux). To unpack this you will need the free gzip (or gunzip) utility available from http://www.gnu.org. 'cd' to the target directory for UMLS2002AA (see below), then type the following command to the file 2002AA_1.TGZ on the CD on your system:

```
gzip -dc [cdrom_path]/2002AA_1.TGZ | tar xvf -
```

where [cdrom_path] is the path to the TGZ file on your CD-ROM.

NOTE that if you do not include the pipe to tar, the extraction will fail when the intermediate file exceeds most Unix file system's 2 GB size limits on single files.

All users should extract the full 2002AA UMLS Knowledge Sources to hard disk, creating the Standard UMLS2002AA Directory Structure below which occupies 3.25 GB. We recommend a minimum of 6-8 GB available disk space.

PLEASE NOTE that you must have the full 2002AA UMLS Knowledge Sources on a local hard disk to use MetamorphoSys; you will need MetamorphoSys to comply with the license agreement and to customize the Metathesaurus to meet your needs.

Standard 2002AA UMLS Knowledge Source Directory Structure

Disc name: 2002AA_TAR or 2002AA_ZIP

/2002AA	root directory
DOC/	UMLS Knowledge Source documentation (this manual) in ASCII and in HTML.
META/	Metathesaurus concepts in ASCII relational format
CHANGE/	Files identifying significant differences from the previous edition.
METAMSYS/	MetamorphoSys system
METASUBSET/	Your customized Metathesaurus (initially empty)
NET/	Semantic Network in ASCII relational format and unit record formats; and Semantic Network documentation in ASCII format.
LEX/	SPECIALIST lexicon in ASCII relational and unit record formats and SPECIALIST documentation in ASCII format.
LEX_PGMS/	SPECIALIST lexicon related lexical programs in executable and C source code.

APPENDIX A

LICENSE AGREEMENT FOR USE OF THE UMLS PRODUCTS

The License Agreement for 2002 is available with the 2002AA Documentation: (license.html) and on the NLM website: (http://www.nlm.nih.gov/research/umls/license.html). NLM does not charge for the UMLS Knowledge Sources. In some circumstances, users of the UMLS Metathesaurus may need to enter into separate license arrangements (See Appendix A.1), which may involve charges, with the copyright holders of some of the individual vocabularies that have been incorporated in the UMLS Metathesaurus.

APPENDIX A.1

Appendix to the License Agreement for Use of the UMLS Knowledge Sources

UMLS METATHESAURUS SOURCE VOCABULARIES -- January 2002AA Edition

Sources are listed in order according to the abbreviations used in the UMLS Metathesaurus files. If additional restrictions and notices apply, the category of restrictions and the special notices appear under the name of the source. See the license agreement for an explanation of the categories of restrictions. Many sources publish printed editions and/or other explanatory information that may be essential to understanding the purpose and application of particular sources in data creation and retrieval. Contact information is provided for each source. Please address questions about permissions or license agreements for additional uses not covered by this Agreement, or other inquiries about individual sources, to the appropriate contacts.

NLM is working toward inclusion in the UMLS Metathesaurus of the complete, current edition of most of these vocabulary sources.

AIR93 AI/RHEUM. Bethesda, (MD): National Library of Medicine, Lister Hill Center, 1993.

Contact: May Cheh, Lister Hill Center, NLM; e-mail: cheh@nlm.nih.gov

---

ALT2000 Alternative Billing Concepts (AltLink). Las Cruces (NM): Alternative Link LLC, 2000.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Alternative Link LLC; 1065 Main Street, Bldg C; Las Cruces, NM 88005; phone: 505-527-0636; http://www.alternativelink.com; e-mail: mail@alternativelink.com

---

AOD99 Alcohol and Other Drug Thesaurus: A Guide to Concepts and Terminology in Substance Abuse and Addiction. 3rd ed. [4 volumes]. Bethesda (MD): National Institute on Alcohol Abuse and Alcoholism (NIAAA) and Center for Substance Abuse Prevention (CSAP), 1999.

Contact: Kathleen Mullen; e-mail: kmullen@his.com

BI98 Beth Israel OMR Clinical Problem List Vocabulary. Version 1.0. Boston (MA): Beth Israel Deaconess Medical Center, 1999.

CATEGORY 2 RESTRICTIONS APPLY

Contact: Howard Goldberg, MD.; e-mail: hgoldber@bidmc.harvard.edu

---

BRMP2002 Descritores em Ciencias da Saude [Portuguese translation of MeSH]. Sao Paulo (Brazil): Latin American and Caribbean Center on Health Sciences Information. BIREME/PAHO/WHO, 2002.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Elenice de Castro; e-mail: elenice@brm.bireme.br

BRMS2002 Descriptores en Ciencias de la Salud [Spanish translation of MeSH]. Sao Paulo (Brazil): Latin American and Caribbean Center on Health Sciences Information. BIREME/PAHO/WHO, 2002.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Elenice de Castro; e-mail: elenice@brm.bireme.br

CCPSS99 Canonical Clinical Problem Statement System (CCPSS). Version 1.0. Nashville (TN): Department of Biomedical Informatics, Vanderbilt University, 1999.

CATEGORY 3 RESTRICTIONS APPLY

Permission will be freely given for any uses and applications containing CCPSS which are not for sale - i.e. those used internally or given to others without charge.

Contact: Steven Brown, MD., Department of Biomedical Informatics, Vanderbilt University; phone: (615)936-1424; e-mail: sbrown@vumclib.mc.vanderbilt.edu

---

CCS99 Clinical Classifications Software (CCS) Categories. June 1999 release. Rockville (MD): Agency for Health Care Policy and Research (AHCPR), 1999.

Contact: phone: 1-800-358-9295; http://www.ahcpr.gov/data/hcup/ccsfact.htm

---

CDT3 Current Dental Terminology (CDT) contained in the HCFA Common Procedure Coding System (HCPCS). Version 3. Washington (DC): Health Care Financing Administration, 2002.

CATEGORY 3 RESTRICTIONS APPLY

For CDT the following special notice must be displayed:
"For CDT only, copyright 2002 American Dental Association, all rights reserved."

Contact for CDT: Mary Essling, American Dental Association; phone: (312)440-2520

COS95, COS93, COS92, COS89 Computer Stored Ambulatory Records (COSTAR). Boston (MA): Massachusetts General Hospital, 1995, 1993, 1992, 1989.

The UMLS Metathesaurus includes terms that were used frequently at 3 COSTAR sites in the years indicated and supplied to NLM by Massachusetts General Hospital.

Contact: G.Octo Barnett, M.D., Laboratory of Computer Science, Massachusetts General Hospital; e-mail: obarnett@warren.med.harvard.edu

---

CPM93 Columbia Presbyterian Medical Center Medical Entities Dictionary (MED). New York (NY): Columbia Presbyterian Medical Center, 1993.

The UMLS Metathesaurus includes a relatively small number of terms created at Columbia Presbyterian Medical Center for the MED, which also includes terms obtained from the UMLS Metathesaurus and other sources.

CATEGORY 2 RESTRICTIONS APPLY.

Contact: James Cimino, M.D.; e-mail: jjc7@columbia.edu;
http://www.cpmc.columbia.edu/homepages/gum7001/topics/index.html

---

CPT01SP Physicians' Current Procedural Terminology (CPT), Spanish translation. 4th ed. Chicago (IL): American Medical Association, 2001.

CATEGORY 3 RESTRICTIONS APPLY.

Contact: http://www.ama-assn.org

CPT2002 Physicians' Current Procedural Terminology (CPT™). 4th ed. Chicago (IL): American Medical Association, 2002.

CATEGORY 3 RESTRICTIONS APPLY.

The following special notice must be displayed:

"CPT™ only Copyright 2002 American Medical Association. All rights Reserved. No fee schedules, basic unit, relative values or related listings are included in CPT™. AMA does not directly or indirectly practice medicine or dispense medical services. AMA assumes no liability for data contained herein.

U.S. Government Rights

This product includes CPT™ which is commercial technical data and/or computer data bases and/or commercial computer software and/or commercial computer software documentation, as applicable which were developed exclusively at private expense by the American Medical Association, 515 North State Street, Chicago, Illinois, 60610. U.S. Government rights to use, modify, reproduce, release, perform, display, or disclose these technical data and/or computer data bases and/or computer software and/or computer software documentation are subject ot the limited rights restrictions of DFARS 252.227-7015(b)(2) (June 1995) and/or subject to the restrictions of DFARS 227.7202-1(a) (June 1995) and DFARS 227.7202-3(a) (June 1995), as applicable for U.S. Department of Defense procurements and the limited rights restrictions of FAR 52.227-14 (June 1987) and/or subject to the restricted rights provisions of FAR 52.227-14 (June 1987) and FAR 52.227-19 (June 1987), as applicable, and any applicable agency FAR Supplements, for non-Department of Defense Federal procurements."

Contact: Shelley J. Jacobs, CPT Intellectual Property Services, American Medical Association, 515 N. State Street, Chicago, IL 60610; fax: (312)464-5762; http://www.ama-assn.org

---

CSP2001 Computer Retrieval of Information on Scientific Projects (CRISP). Bethesda (MD): National Institutes of Health, Division of Research Grants, Research Documentation Section, 2001.

Contact: Dorrette Finch Worrell, Chief, Research Documentation Section, NIH; e-mail: dw33v@nih.gov; http://www-commons.cit.nih.gov/crisp

---

CST95 Coding Symbols for Thesaurus of Adverse Reaction Terms (COSTART) . Rockville (MD): Food and Drug Administration, Center for Drug Evaluation and Research, 1995.

COSTART will be superseded by the Medical Dictionary for Regulatory Activities (MedDRA) Terminology.

Contact: Available from the National Technical Information Service
Website: http://www.ntis.gov/fcpc/cpn5580.htm

DDB00 Diseases Database 2000. London (England): Medical Object Oriented Software Enterprises Ltd., 2000.

CATEGORY 3 RESTRICTIONS APPLY.

Contact: Malcolm Duncan; http://www.diseasesdatabase.com/

DMD2002 German translation of the MeSH. Cologne (Germany): Deutsches Institut fuer Medizinische Dokumentation und Information, 2002.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Dr. Elisabeth Berg-Schorn; e-mail: helpdesk@dimdi.de; http://www.dimdi.de

---

DMDICD Internationale Klassifikation der Krankheiten 10 [German translation of ICD10]. Cologne (Germany): Deutsches Institut fuer Medizinische Dokumentation und Information, 2000.

CATEGORY 1 RESTRICTIONS APPLY

Contact: Dr. Michael Schopen; e-mail: schopen@dimdi.de; http://www.dimdi.de

---

DMDUMD Die Nomenklatur fuer Medizinprodukte UMDNS [German translation of UMDNS]. Cologne (Germany): Deutsches Insitut fuer Medizinische Dokumentation und Information, 2000.

CATEGORY 1 RESTRICTIONS APPLY

Contact: Dr. Michael Schopen; e-mail: schopen@dimdi.de; http://www.dimdi.de

---

DSM3R Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R). 3rd ed. rev. Washington (DC): American Psychiatric Association, 1987.

CATEGORY 3 RESTRICTIONS APPLY

Contact: see entry for DSM4

DSM4 Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Washington (DC): American Psychiatric Association, 1994.

CATEGORY 3 RESTRICTIONS APPLY

The APA usually charges small administrative fees for copyright permissions, but these may be waived for research purposes. All users should apply for permission in writing or by email to:

Contact: Customer Service, American Psychiatric Press, Inc. 1400 K Street, N.W. Washington, DC 2005; e-mail: csdept@appi.org; http://www.appi.org

---

DUT2001 Nederlandse vertaling van Mesh (Dutch translation of MeSH). Amsterdam, The Netherlands: Nederlands Tijdschrift voor Geneeskunde (Dutch Journal of Medicine), 2001.

CATEGORY 3 RESTRICTIONS APPLY

Contact: A.J.P.M.Overbeke, overbeke@ntvg.nl; Nederlands Tijdschrift voor Geneeskunde (Dutch Journal of Medicine), P.O.Box 75971, 1070 AZ Amsterdam, The Netherlands; http://www.ntvg.nl

---

DXP94 DXplain (An expert diagnosis program). Boston (MA): Massachusetts General Hospital.

Contact: G. Octo Barnett, M.D., Laboratory of Computer Science, Massachusetts General Hospital; e-mail: obarnett@partners.org; http://www.lcs.mgh.harvard.edu

---

FIN2002 Finnish translation of MeSH. Helsinki (Finland): Finnish Medical Society Duodecim, 2002.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Peter Nyberg, MD; e-mail: peter.nyberg@sll.fimnet.fi

---

HCDT3 HCPCS Version of Current Dental Terminology (CDT). Version 3. Washington (DC): Health Care Financing Administration, 2002.

CATEGORY 3 RESTRICTIONS APPLY

HCPCS02 Health Care Financing Administration (HCFA) Common Procedure Coding System (HCPCS). Washington (DC): Health Care Financing Administration, 2002.

The American Medical Association's CPT™ codes in HCPCS have a Source Abbreviation of HCPT02. The American Dental Association's CDT codes in HCPCS have a Source Abbreviation of HCDT3.

Contact for HCPCS: C. Kaye Riley, Health Care Financing Administration; e-mail: criley@hcfa.gov

---

HCPT02 HCPCS Version of Current Procedural Terminology (CPT). Washington (DC): Health Care Financing Administration, 2002.

CATEGORY 3 RESTRICTIONS APPLY

HHC96 Saba, Virginia. Home Health Care Classification of Nursing Diagnoses and Interventions. Washington (DC): Georgetown University, 1996.

CATEGORY 1 RESTRICTIONS APPLY

Contact: Virginia Saba, Ph.D., R.N.,; e-mail: vsaba@worldnet.att.net; phone: (703)521-6132; fax: (703)521-3866; http://www.sabacare.com/

HL7 Health Level Seven Vocabulary (HL7). Ann Arbor (MI): Health Level Seven, 1998.

Contact: Health Level Seven, 3300 Washtenaw Avenue, Suite 227, Ann Arbor MI 48104-4250; phone: (734)677-7777; fax: (734)677-6622; e-mail: HQ@HL7.ORG; http://www.hl7.org

---

ICD10, ICD10AE International Statistical Classification of Diseases and Related Health Problems (ICD-10). Tenth Revision. Geneva (Switzerland): World Health Organization, 1998.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Office of Publications, World Health Organization, 1211 Geneva 27, Switzerland; http://www.who.int/whosis/icd10

ICD10AM International Statistical Classification of Diseases and Related Health Problems. Tenth Revision, Australian Modification; 2nd Edition: January 2000.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Developed and Maintained by the National Centre for Classification in Health, University of Sydney, Faculty of Health Sciences; PO Box 170 Lidcombe, NSW, Australia 1825; Phone: 2 9351 9461

ICD2002 International Classification of Diseases: 9th revision, Clinical Modification (ICD-9-CM). 6th ed. Washington (DC): Health Care Financing Administration; July 2001.

NLM has generated fully specified titles for ICD-9-CM codes in cases in which the official ICD- 9-CM titles consist of extensions to higher levels in the ICD-9-CM hierarchy. The fully specified names were produced with reasonable care, but have not yet been reviewed and approved by the producers of ICD-9-CM.

Contact for Diseases: Donna Pickett, National Center for Health Statistics; e-mail: dfp4@cdc.gov; http://www.cdc.gov/nchs

Contact for Procedures: Patricia Brooks, Health Care Financing Administration; e-mail: pbrooks@hcfa.gov; http://www.hcfa.gov

---

ICPC2E, ICPC2AE International Classification of Primary Care: 2nd ed. electronic form. Classification Committee of the World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians (WONCA), known more briefly as the World Organization of Family Doctors; 1998.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Henk Lamberts (H.Lamberts@AMC.UVA.NL) and Inge Okkes (I.m.okkes@amc.una.nl)

See reference: Okkes, IM; Jamoulle, M; Lamberts, H; Bentzen, N. ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences. *Family Practice*; 2000; 17:101-107.

ICPC2P, ICPCPAE International Classification of Primary Care: Version 2-Plus: Australian Modification; January, 2000.

CATEGORY 3 RESTRICTIONS APPLY

ICPC93 Lamberts, Henk, Wood, Maurice, and Hofmans-Okkes, Inge, editors. International Classification of Primary Care (ICPC) in The European Community (with a multi language layer on a floppy disk). Oxford: Oxford University Press, 1993.

This year, the Metathesaurus has also included translations of ICPC93 in Basque (ICPCBAQ), Danish (ICPCDAN), Dutch (ICPCDUT), Finnish (ICPCFIN), French (ICPCFRE), German (ICPCGER), Hebrew (ICPCHEB), Hungarian (ICPCHUN), Italian (ICPCITA), Norwegian (ICPCNOR), Portuguese (ICPCPOR), Spanish (ICPCSPA), and Swedish (ICPCSWE).

Contact: Henk Lamberts; e-mail: H.Lamberts@AMC.UVA.NL or Inge Okkes; e-mail: I.m.okkes@amc.una.nl; University of Amsterdam

---

INS2001 Thesaurus Biomedical Francais/Anglais [French translation of MeSH]. Paris (France): Institut National de la Sante et Recherche Medicale, 2001.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Dr. Annie Advocat; e-mail: advocat@inserm-dicdoc.u-strasbg.fr; http://www.inserm.fr

---

ITA2002 Italian translation of MeSH. Rome (Italy): Instituto Superiore di Sanita Servizio Documentazione, 2002.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Dr. Adriana Dracos, Viale Regina Elena, 229 00616 Rome Italy; phone: 39-06-49902280; fax: 39-06-49387117; e-mail: dracos@iss.it

---

JABL99 Online Congenital Multiple Anomaly/Mental Retardation Syndromes. Stanley Jablonski. Bethesda (MD): National Library of Medicine, 1999.

CATEGORY 1 RESTRICTIONS APPLY

Contact: Stanley Jablonski, National Library of Medicine, Bethesda MD; e-mail: stanley_jablonski@nlm.nih.gov; http://www.nlm.nih.gov/mesh/jablonski/syndrome_title.html

---

LCH90 Library of Congress Subject Headings. 12th ed. Washington (DC): Library of Congress, 1989.

There are later editions of this source that are not reflected in the UMLS Metathesaurus. This source has considerable non-biomedical content and will never be included in the Metathesaurus in its entirety.

Contact: http://www.lcweb.loc.gov

LNC203 Logical Observations Identifiers, Names, and Codes (LOINC). Version 2.03. Indianapolis (IN): The Regenstrief Institute, 2001.

Contact: http://www.regenstrief.org/loinc/loinc.htm

MCM92 Haynes, Brian. Glossary of Methodologic Terms for Clinical Epidemiologic Studies of Human Disorders. Hamilton, Ontario (Canada): McMaster University, 1992.

Contact: R. Brian Haynes, M.D., Ph.D.; e-mail: bhaynes@fhs.mcmaster.ca

---

MDDB99 First DataBank Master Drug Data Base (MDDB). San Bruno (CA): First DataBank Inc., 1999.

CATEGORY 3 RESTRICTIONS APPLY

Contact: http://www.firstdatabank.com/

MDR40 Medical Dictionary for Regulatory Activities Terminology (MedDRA) Version 4.0, June, 2001. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

CATEGORY 3 RESTRICTIONS APPLY

Contact: http://meddramsso.com

MDRAE40 Medical Dictionary for Regulatory Activities Terminology (MedDRA), American English Equivalents, Version 4.0, June, 2001. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

CATEGORY 3 RESTRICTIONS APPLY

Contact: http://meddramsso.com

MDREA40 Medical Dictionary for Regulatory Activities Terminology (MedDRA), American English, with expanded abbreviations, Version 4.0, June, 2001. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

CATEGORY 3 RESTRICTIONS APPLY

Contact: http://meddramsso.com

MDREX40 Medical Dictionary for Regulatory Activities Terminology (MedDRA),with expanded abbreviations, Version 4.0, June, 2001. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

CATEGORY 3 RESTRICTIONS APPLY

Contact: http://meddramsso.com

MIM93 Victor A. McKusick, ed. Online Mendelian Inheritance in Man (OMIM). Baltimore (MD): Johns Hopkins University, 1994.

To date the UMLS Metathesaurus contains a relatively small amount of data from this source.

CATEGORY 1 RESTRICTIONS APPLY

Contact: Joanna Amberger, Johns Hopkins University; e-mail: joanna@ncbi.nlm.nih.gov; http://www.ncbi.nlm.nih.gov/omim (This site contains the entire database).

MMSL01 Multum MediSource Lexicon. Denver (CO): Multum Information Services, Inc., July 2001.

CATEGORY 1 RESTRICTIONS APPLY

Contact: 1-888-633-4772; http://www.multum.com

MMX01 Micromedex DRUGDEX. Englewood (CO): Micromedex, 2001.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Micromedex, 6200 South Syracuse Way, Suite 300, Englewood CO 80111-4740; phone: (800) 525-9083; e-mail: info@mdx.com; http://www.micromedex.com

---

MSH2002 Medical Subject Headings (MeSH). Bethesda (MD): National Library of Medicine, 2002.

This source has been translated into many languages. To date, eight of the translations have been incorporated into the UMLS Metathesaurus.

Contact: Stuart Nelson, M.D., Head, MeSH Section; e-mail: nelson@nlm.nih.gov; http://www.nlm.nih.gov/mesh/meshhome.html

---

MTH UMLS Metathesaurus. Bethesda (MD): National Library of Medicine.

Concept names with this source abbreviation were created by NLM to facilitate creation of the UMLS Metathesaurus. There are relatively few of them.

Contact: William T. Hole, M.D., Lister Hill Center, NLM; e-mail: wth@nlm.nih.gov

---

MTHCH02 UMLS Metathesaurus Hierarchical CPT terms. Bethesda (MD): National Library of Medicine.

Concept names with this source abbreviation were created by NLM to provide contextual information for CPT™.

CATEGORY 3 RESTRICTIONS APPLY

Contact: William T. Hole, M.D., Lister Hill Center, NLM; e-mail: wth@nlm.nih.gov

---

MTHHH02 Metathesaurus Hierarchical HCPCS Terms (These terms were created by the NLM to provide contextual information for HCPCS). Bethesda (MD): National Library of Medicine.

Concept names with this source abbreviation were created by NLM to provide contextual information for HCPCS.

Contact: William T. Hole, M.D., Lister Hill Center, NLM; e-mail: wth@nlm.nih.gov

---

MTHICD9 NLM-generated Entry Terms for ICD-9-CM. Bethesda (MD): National Library of Medicine.

Contact: William T. Hole, M.D., Lister Hill Center, NLM; e-mail: wth@nlm.nih.gov

---

MTHMST2001 Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy: International Edition April 22, 1998.

Contact: Michele Tringali, tringali.michele@aoud.sanita.fvg.it

---

MTHMSTFRE Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy: French Edition April 22, 1998.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Michele Tringali, tringali.michele@aoud.sanita.fvg.it

---

MTHMSTITA Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy: Italian Edition April 22, 1998.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Michele Tringali, tringali.michele@aoud.sanita.fvg.it

---

NAN99 Carroll-Johnson, Rose Mary, editor. Classification of Nursing Diagnoses. Proceedings of the 10th conference, North American Diagnosis Association. Philadelphia (PA): Lippincott, 1999.

CATEGORY 3 RESTRICTIONS APPLY to commercial use.

Contact: NANDA, 1211 Locust St, Philadelphia, PA 19107; phone: (800)647-9002

NCI2001a NCI Thesaurus. Version 3.0. Bethesda (MD): National Cancer Institute, National Institutes of Health, July 2001.

Contact: Sherri de Coronado; e-mail: decorons@exchange.nih.gov

---

NDDF01 National Drug Data File. San Bruno (CA): First DataBank Inc., June 2001.

CATEGORY 3 RESTRICTIONS APPLY.

Contact: http://www.firstdatabank.com/

NEU99 Bowden, Douglas M., Martin, Richard F., Dubach, Joev G. Neuronames Brain Hierarchy. Seattle (WA): University of Washington, Primate Information Center, 1999.

CATEGORY 3 RESTRICTIONS APPLY.

Contact: Douglas M. Bowden, M.D., Regional Primate Research Center, University of Washington, Box 357330, Seattle, WA 98195; e-mail: dmbowden@u.washington.edu; http://rprcsgi.rprc.washington.edu/neuronames/index.html

---

NIC99 McCloskey, Joanne C.; Bulechek, Gloria M., editors. Nursing interventions classification (NIC): Iowa intervention project. St. Louis (MO): Mosby Year Book, 1999.

CATEGORY 3 RESTRICTIONS APPLY.

Contact: Katherine Taylor, Mosby-Year Book; phone: (800)345-8738, ext. 7419; e-mail: katherine.taylor@mosby.com; http://www.nursing.uiowa.edu/nic

---

NOC97 Johnson, Marion; Maas, Meridean, editors. Nursing Outcomes Classification (NOC): Iowa outcomes project. St. Louis (MO): Mosby Year Book, 1997.

CATEGORY 3 RESTRICTIONS APPLY.

Contact: Katherine Taylor, Mosby-Year Book; phone: (800)345-8738, ext. 7419; e-mail: katherine.taylor@mosby.com; http://www.nursing.uiowa.edu/noc

---

OMS94 Martin, Karen S., Scheet, Nancy J. The Omaha System: Applications for Community Health Nursing. Philadelphia (PA): W.B. Saunders, 1992. (with 1994 corrections)

CATEGORY 1 RESTRICTIONS APPLY.

Contact: Karen S. Martin, RN, MSN, FAAN, 2115 South 130th Street, Omaha, Nebraska 68144; fax: (402)333-2091

PCDS97 Ozbolt, Judy Grace. Patient Care Data Set (PCDS). Version 4.0. Nashville (TN): Vanderbilt University School of Nursing, 1998.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Judy Ozbolt, Vanderbilt School of Nursing, 400-C Godchaux Hall, Nashville TN 37240-0008; phone: (615)343-3291; e-mail: judy.ozbolt@mcmail.vanderbilt.edu

---

PDQ2001 Physician Data Query Online System (PDQ). Bethesda (MD): National Cancer Institute, July 2001.

Contact: Dr. Gisele Sarosy, National Cancer Institute; e-mail: gisele@icic.nci.nih.gov; http://cancernet.nci.nih.gov

---

PPAC98 Pharmacy Practice Activity Classification (PPAC). Washington (DC): American Pharmaceutical Association. January, 1998.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Scott Antall, American Pharmaceutical Association - Academy of Pharmaceutical Research and Science; e-mail: ssa@mail.aphanet.org

---

PSY2001 Thesaurus of Psychological Index Terms. Ninth Edition. Washington (DC): American Psychological Association, 2001.

CATEGORY 3 RESTRICTIONS APPLY.

Contact: PsycINFO Permissions; phone: (800)374-2722; e-mail: PsycINFO@APA.org; http://www.apa.org

---

QMR96 Quick Medical Reference (QMR). San Bruno (CA): First Databank, 1997.

Contact: http://www.firstdatabank.com

RAM99 Randolph A. Miller Clinically Related Concepts. Nashville (TN): Vanderbilt University, 1999.

Contact: Randy Miller, randolph.a.miller@vanderbilt.edu

RCD99, RCDAE, RCDSA, RCDSY

Clinical Terms Version 3 (Read Codes) (Q199). England: National Health Service Centre for Coding and Classification, March 1999.

CATEGORY 3 RESTRICTIONS APPLY

Contact: NHS Information Authority Loughborough, Woodgate, Loughborough, Leicestershire LE11 2TG; fax: (0) 1509 211611; e-mail: helpdesk3@nhsccc.exec.nhs.uk

---

RUS2002 Russian Translation of MeSH. Moscow (Russia): State Central Scientific Medical Library, 2002.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Natalya Lomova, Head of Department State Central Scientific Medical Library, Nakhimovsky prospekt 49, Moscow, GSP 117418, Russia; e-mail: vld@vyugin.mccme.rssi.ru

---

SNM2 Cote, Roger A., editor. Systematized nomenclature of medicine. 2nd ed. Skokie (IL): College of American Pathologists, 1979. SNOMED update, 1982. Skokie (IL): College of American Pathologists, 1982.

CATEGORY 3 RESTRICTIONS APPLY

Contact: See entry for SNMI.

SNMI98 Cote, Roger A., editor. Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International. Version 3.5. Northfield (IL): College of American Pathologists; Schaumburg (IL): American Veterinary Medical Association, 1998.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Jill Rosenthal/Naomi Siebert, College of American Pathologists, 325 Waukegan Road, Northfield, IL
60093-2750; fax: (847)832-8335; e-mail: jnrs@cap.org; http://www.snomed.org

---

SPN99 Standard Product Nomenclature (SPN). Rockville (MD): U.S. Food and Drug Administration, 1999.

Contact: Ann Tornese; e-mail: act@fdadr.cdrh.fda.gov

SRC Entries for UMLS Metathesaurus Source Terminologies. Bethesda (MD): National Library of Medicine.

Contact: William T. Hole, M.D., Lister Hill Center, NLM; e-mail: wth@nlm.nih.gov

---

ULT93 Bell, Douglas. Ultrasound Structured Attribute Reporting (UltraSTAR). Boston (MA): Brigham & Womens Hospital, 1993.

CATEGORY 3 RESTRICTIONS APPLY

Contact: Robert Greenes, M.D., Ph.D., Brigham & Womens Hospital; e-mail: greenes@harvard.edu

---

UMD2002 Universal Medical Device Nomenclature System: Product Category Thesaurus. Plymouth Meeting (PA): ECRI, 2002.

CATEGORY 1 RESTRICTIONS APPLY.

Contact: Elizabeth Richardson, Director of Database and Nomenclature Systems, ECRI; e-mail: erichard@ecri.org; http://www.ecri.org

---

UWDA155 University of Washington Digital Anatomist Symbolic Knowledge Base. Rosse Cornelius. Structural Informatics Group, Department of Biological Structure, University of Washington, Seattle, Version 1.5.5, January 2001.

Contact: Jose Mejino, M.D.; e-mail: onard@biostr.washington.edu; http://sig.biostr.washington.edu/projects/da/

VANDF01 U.S. Department of Veterans Affairs, Veterans Health Administration National Drug File. Department of Veterans Affairs, Washington, DC. Release Date: September 5, 2001.

CATEGORY 3 RESTRICTIONS APPLY

Contact: http://www.vapbm.org/PBM/natform1.htm

WHO97 WHO Adverse Drug Reaction Terminology (WHOART). Uppsala (Sweden): WHO Collaborating Centre for International Drug Monitoring, 1997.

WHOART will be superseded by the Medical Dictionary for Regulatory Activities (MedDRA) Terminology.

CATEGORY 2 RESTRICTIONS APPLY.

The Metathesaurus includes translations of WHO97 in French (WHOFRE), German (WHOGER), Portuguese (WHOPOR), and Spanish (WHOSPA).

Contact: WHO Collaborating Centre for International Drug Monitoring, Stora Target 3, S-753 20
Uppsala, Sweden; fax: 18-656080; http://www.who.pharmasoft.se/prodserv.html

APPENDIX B

METATHESAURUS DATA ELEMENTS

AND SOURCE VOCABULARY INFORMATION

B.0 Introduction

This appendix contains descriptions of the Metathesaurus MR* file columns (Section B.1.1), Metathesaurus Attribute values (Section B.1.2), a listing of the vocabularies and classifications that are the sources of concepts and terms in the Metathesaurus with the source abbreviations used in the Metathesaurus data (Section B.2 and B.2.1), a listing of the number of strings from each source (Section B.3), a listing of the types of names that appear in the source vocabularies with the type abbreviations used in the Metathesaurus, e.g., preferred term, entry term (Section B.4), the order of precedence of sources and types of concept names used to determine the preferred name in the Metathesaurus (Section B.5), and a list of all the source specific relationship attributes that are not in the Semantic Network (Section B.6). Appendix B provides supporting details for Section 2 of the documentation. Section 2 contains many references to this Appendix.

B.1 Metathesaurus Data Elements

B.1.0 Introduction

All data elements in the Metathesaurus are described in this section. The data elements have been divided into Column Descriptions and Attribute Descriptions. The descriptions are arranged alphabetically by data element abbreviation.

The designation "Repeating Element" indicates that the data element may have multiple values for a single concept, term, or string.

B.1.1 Column Descriptions

ATN Attribute name

2 or 3 characters

Found in MRSAT

All attribute names are listed in Appendix B.1.2 with descriptions given of each.

Example:

HAB
RN
CPA

ATV Attribute value

Variable length alphanumeric string with embedded punctuation

Found in MRSAT

Attribute values are described under the corresponding attribute name in Section B.1.2. Some attribute values can exceed 3,500 characters.

ATX Associated Expression

Multi-term expression for concept's meaning in a named vocabulary

Found in MRATX

Several different formats of the expression are possible:

a.) <Main Heading>/<Subheading> or

<Main Heading>/<Subheading>/<Subheading> or

b.) Individual Main Headings and/or Main Heading and subheading combinations coordinated by the Boolean operators, (i.e., AND, OR, AND NOT). The vocabulary terms in the expression are included in less than and greater than signs <> to distinguish them from Boolean operators. Parentheses are used as necessary to clarify the meaning of Boolean OR expressions.

Examples:

<Pupil>/<Abnormalities>

<Polyps> AND <Rectal Neoplasms>

(<Rectum> OR <Rectal Diseases>) AND <Pain>

Note: Users' needs for Associated Expressions and more formalized data structure are under review. Please provide feedback if you find them useful and/or have suggestions, since we must justify the cost and effort of further development and maintenance.

An expression consists of multiple terms in one Metathesaurus source vocabulary which best expresses the meaning of a concept. In the 2002AA Metathesaurus, these expressions are constructed of Medical Subject Headings (MeSH); of Library of

Congress Subject Headings (LCSH); or of Omaha System terms. Information in this element can be used to construct a MeSH or LCSH search related to the concept represented in the Metathesaurus entry, or to determine the correct relationships between concept/qualifier combinations in the Omaha System and concepts that are narrower than these combinations.

Note that parentheses are used in these expressions to clarify the Boolean expression, and punctuation is used to distinguish the vocabulary terms from the Boolean operators. The punctuation will probably have to be removed before the expressions can be used in a search. Depending on the system being searched, the expression may have to be divided into multiple search statements to achieve the desired search result. In the case of MeSH expressions, it may be necessary to EXPLODE one or more terms in the expression to get the desired search result. In the case of LCSH expressions, some of the subheadings are followed by "etc." indicating that other similar subheadings may be used. Note that the LCSH expressions represent combinations actually used by the Library of Congress in cataloging books. They have been mapped to their closest MeSH equivalents.

AV Average length

Numeric value

Found in MRCOLS

Describes the average length of data in a column.

BTS Size in bytes

Numeric value

Found in MRFILES

For each MR* file, lists the size in bytes.

CLS Number of columns

Numeric value

Found in MRFILES

For each MR* file, lists the # of columns present.

COA Attributes of co-occurrence relationship

Found in MRCOC

For MED02 and MBD02 source only

Subheading data for co-occurring concept. Multiple values separated by commas.

Each value contains the two letter abbreviation of a MeSH subheading applied to first concept when it co-occurs with the second concept or <> (indicating no subheading), followed by the frequency of application of this subheading to first concept when it co-occurs with the second.

The 2-letter abbreviations for subheadings - e.g. AB = "abnormalities" - may be found in the MRSAT "QA" attribute name for the subheading concepts, or in the MeSH documentation, e.g. the MeSH browser at:

http://nls10.nlm.nih.gov:1082/MBrowser.html

COF Frequency of Co-occurrence

Found in MRCOC

The frequency of co-occurrence is the count of CUI1 with CUI2 co-occurrence. In a single citation, many MeSH headings may have more than one subheading applied.

COT Type of Co-occurrence

Found in MRCOC

Valid values for Type of Co-occurrence:

L	Co-occurrence of primary or main subject headings in citations to the published literature
LQ	second concept occurs as a MeSH topical qualifier of the first in citations to the published literature. Where CUI2 is not present, the count of citations of CUI1 with no MeSH qualifiers is reported.
LQB	second concept is qualified by the first (a MeSH topical qualifier) in citations to the published literature
KP	positive association in Knowledge Base
KN	negative association in Knowledge Base, e.g., a finding that is inconsistent with a disease.
MP	Co-occurrence of modifier and problem within a patient record
PP	Co-occurrence of two problems within a patient record

CREL Relationship of retired CUI (CUI1) to current CUI (CUI2)

Found in MRCUI

Allowed values:

DEL - deleted - no related CUI present
SY - merged - the two CUIs are synonymous
RN - Narrower - CUI2 is narrower than CUI1
RB - Broader - CUI2 is broader than CUI1
RO - Other - CUI2 has other relationship to CUI1

CUI Unique Identifier for Concept

The letter C followed by seven digits

The unique identifier for the Metathesaurus concept to which a term and string are linked. The identifier remains the same across versions of the Metathesaurus irrespective of the term designated as the preferred name of the concept. This facilitates file maintenance and management. In the Metathesaurus distribution formats, all data elements are linked to the concept(s) to which they belong.

Note: See MRCUI for cases where CUIs were retired.

CUI1 Unique Identifier of first concept

Letter C followed by 7 numbers

Found in MRREL, MRCOC, MRCUI, and MERGED.CUI

CUI2 Unique Identifier of second concept

Letter C followed by 7 numbers

Found in MRREL, MRCOC, MRCUI, and MERGED.CUI

CXL Context member label

3 letter abbreviation

Found in MRCXT

Possible values are ANC for ancestor of this concept, CCP for concept, SIB for sibling of this concept, and CHD for child of this concept.

CXN Context number

Number character

Found in MRCXT

It is used to distinguish multiple contexts in the same source with the same String Unique Identifier (SUI). The CXNs are not global but are created as required for each concept. Since some concepts have multiple contexts in the same vocabulary with the same SUI, each distinct context can be retrieved with a CUI-SUI-SAB-CXN key.

CXS String for context member

Alphanumeric value

Found in MRCXT

CXT Context or hierarchy

Repeating element

Found in MRCXT

Contexts, hierarchical or otherwise, for the concept from various Metathesaurus source vocabularies. The contexts are available for display to users as an aid to understanding the scope of concepts and for refinement of search strategies. In most cases, the "descendants" of a concept in one or more hierarchies can be usefully assembled to enhance the search for that concept.

The contexts for the Omaha System (OMS) include qualifiers and may require special handling. In the Omaha System, the same qualifier, e.g., Deficit, may appear as the second level in multiple contexts. Its descendants in each context are uniquely applicable to a specific concept/qualifier combination. It is therefore inappropriate to gather all of its descendants to expand a single search.

DEF Definition

Repeating element

Found in MRDEF

Two subelements

1. Abbreviation of the source of the definition (see Section B.2 for valid values)
2. Variable length string with alpha characters, punctuation, and in some cases numerics

Narrative description(s) of the meaning of the concept. The majority of the definitions come from MeSH, but there are also definitions from a number of other sources. The definitions created specifically for the Metathesaurus have a source value of MTH. Definitions were created specifically for the Metathesaurus only when needed to distinguish among different meanings of the same string. Note that some definitions may be several thousand characters long and tab characters are used to indicate line structure within a few definitions. See also SOS Scope Statement.

DES Descriptive name

Alphabetic value with embedded punctuation

Found in MRFILES and MRCOLS

DTY Suggested SQL-92 data type

Alphanumeric value

Found in MRCOLS

SQL-92 data type for this column.

FIL Physical FILENAME

Alphabetic value

Found in MRFILES and MRCOLS

Valid values are the MR* file names.

FMT List of Columns

Alphabetic value

Found in MRFILES

It is a comma separated list of all the columns in a file in order.

FR Frequency count of number of occurrences of a concept in the information source

Numeric value

Found in MRLO

Optional field.

HCD Hierarchical number or code of context member

Alphanumeric value

Found in MRCXT

Optional field.

ISN Name of information source or database

Alphanumeric value

Found in MRLO

Valid Values for Name of information source:

HDA99	Health Devices Alerts (ECRI) (1999)
HPC99	Health Product Comparison System (ECRI) (1999)
MBD02/MED02	MEDLINE (National Library of Medicine) will be followed by an indication of applicable range of years of MEDLINE data
DXP94	DXPLAIN a diagnostic prompting system (Massachusetts General Hospital (1996)
OMIM97	OMIM Online Mendelian Inheritance in Man (Victor McKusick, Johns Hopkins University) (1996)
PDQ2001	PDQ Physician Data Query System (National Cancer Institute) (2001)
QMR97	QMR Quick Medical Reference (First Databank) (1996)
AIR93	AI/RHEUM (National Library of Medicine) (1996)

LAT Language of Term

Three alpha characters

Found in MRCON, the Word Index files and DELETED.SUI

Valid values:

BAQ	Basque
DAN	Danish
DUT	Dutch
ENG	English
FIN	Finnish
FRE	French
GER	German
HEB	Hebrew
HUN	Hungarian
ITA	Italian
NOR	Norwegian
POR	Portuguese
RUS	Russian
SPA	Spanish
SWE	Swedish

The USMARC (Machine-Readable Cataloging) abbreviation for the language of the term. Present in the 2002AA Metathesaurus are Basque, Danish, Dutch, English, Finnish, French, German, Hebrew, Hungarian, Italian, Norwegian, Portuguese, Spanish, and Swedish terms.

LRL Least Restriction Level

Integer

Found in MRCON

The least restrictive "Category of additional restrictions" of the vocabularies which are sources for this string. See SRL for the meaning of each level and the UMLS License agreement for the restriction level of any source vocabulary.

LUI Lexical Unique Identifier

The letter L followed by seven digits

Found in MRCON, MRSAT, MRSO, Word Indexes, AMBIG.LUI, DELETED.LUI and MERGED.LUI

The unique identifier for the group of lexical variants (term) of which a string is one form. A LUI may be linked to multiple CUIs in the Metathesaurus. Strings which share a LUI have the same normalized string form, as produced by the Specialist "norm" program distributed with the Metathesaurus. Changes in "norm" versions may cause strings to be assigned to different LUIs in different editions of the Metathesaurus.

MAX Maximum Length, in bytes

Numeric value

Found in MRCOLS

MG Imprecise Relationship Indicator

Found in MRREL

Valid Value for Imprecise Relationship Indicator:

The single character G.

This indicator is used when the Relationship Labels have been machine-generated for the Metathesaurus based on ambiguous information from a source and have not been reviewed again in the Metathesaurus. There is therefore a possibility that the nature of the relationship has been labeled incorrectly. In the current version of the Metathesaurus, there are no relationships with this indicator. It was used in previous editions for MeSH Supplementary Concepts which had not been reviewed in a concept-oriented system.

MIN Minimum Length, in bytes

Numeric value

Found in MRCOLS

NSTR Normalized String

Alphanumeric value

Found in MRXNS.ENG

In lowercase form.

NWD Normalized Word

Alphanumeric value

Found in MRXNW.ENG

In lowercase form.

RANK Termgroup ranking

Numeric value

Found in MRRANK

Is a numeric order of precedence with the higher value winning.

Example: 0210, 0209

REF Documentation Section Number

Numeric value

Found in MRCOLS

Currently empty because all are linked from the file description to this appendix.

REL Relationship

Found in MRREL and MRATX.

The relationship which the second concept (with Concept Unique Identifier CUI2) HAS TO the first concept (with Concept Unique Identifier CUI1).

RELs may be derived from a source vocabulary's explicit hierarchy (see also MRCXT), derived from other relationships in a source vocabulary, created from information about allowed qualifiers in a source vocabulary, found in Metathesaurus QA of lexical and semantic matches, or added by Metathesaurus editors.

There are separate RELS for each direction of the relationship, e.g., one entry for "Atrial Fibrillation" as a child of "Arrhythmia" and another entry for "Arrhythmia" as a parent of "Atrial Fibrillation".

*Note: The allowed values for the REL field differ in MRREL and MRATX.

Valid Values for REL in MRREL:

RB	has a broader relationship
RN	has a narrower relationship
RO	has relationship other than synonymous, narrower, or broader
RL	the relationship is similar or "alike". the two concepts are similar or "alike". In the current edition of the Metathesaurus, most relationships with this attribute are mappings provided by a source, named in SAB and SL; hence concepts linked by this relationship may be synonymous, i.e. self-referential: CUI1 = CUI2. In previous releases, some MeSH Supplementary Concept relationships were represented in this way.
RQ	related and possibly synonymous.
SY	source asserted synonymy.
PAR	has parent relationship in a Metathesaurus source vocabulary
CHD	has child relationship in a Metathesaurus source vocabulary
SIB	has sibling relationship in a Metathesaurus source vocabulary.
AQ	is an allowed qualifier for the first concept in a Metathesaurus source vocabulary.
QB	can be qualified by.

Valid Values for REL in MRATX:

RB	has a broader relationship
RN	has a narrower relationship
RO	has relationship other than synonymous, narrower, or broader
SY	has synonymous relationship
RU	has an unspecified relationship

RELA Relationship Attribute

An attribute characterizing the nature of a relationship

Found in MRREL and MRCXT

Valid Values for Relationship Attributes:

Any of the relationships defined in the UMLS Semantic Network; these relationships are more principled; examples are isa, part_of, branch_of, result_of;

or

a more specific relationship provided by the source identified in the SAB (Abbreviation for source vocabulary) which is not one of the Semantic Network Relationships. These RELAs are listed in Appendix B.6. Examples are mapped_to, clinically_associated_with, classified_as. Both the SAB and RELA should be considered in deciding to use or reject relationships;

or

no value where further specification is not available.

The large majority of relationships have no Relationship Attribute, although the number with such attributes is increasing and this is a focus of NLM development.

RNK Rank

Numeric value

Found in MRCXT

For rows with a CXL value of ANC, the rank of the ancestors (e.g., a value of 1 denotes the most remote ancestor in the hierarchy).

RWS Number of Rows

Numeric value

Found in MRFILES

SAB Source Abbreviation for source vocabulary

Up to 7 characters

Found in MRREL, MRSAT, MRDEF, MRSO, MRCXT, MRATX, and MRRANK

An abbreviation of the name of a Metathesaurus source vocabulary. In some cases, the year is also indicated. Allowed values are listed in Appendix B, Section B.2, Vocabulary Source Abbreviations and in the SABDOC Bonus file available at:

http://umlsinfo.nlm.nih.gov

SCD Unique identifier or code for an entry in the source vocabulary

Alphanumeric value

Found in MRSAT, MRSO and MRCXT

In cases where there is no native identifier, SCD is a unique identifier generated for the Metathesaurus.

SL Source of Relationship labels

Found in MRREL.

Valid Values for Source of Relationship Labels:

Any source vocabulary abbreviation (see Section B.2).

This is the source of the information about the specific nature of the relationship, i.e., RB, RN, or RO and any Relationship attribute (RELA) values. In many cases, this is the same as the source of the relationship. In some cases, the particular nature of the relationship was added in Metathesaurus construction and the Source of Relationship Labels is MTH. Note that SL is not simply the source of the RELA.

SNA Actual name that occurs in the information source if not otherwise present in the Metathesaurus

Alphanumeric value

Found in MRLO

An optional field.

SOC Source of co-occurrence information in MRCOC

Alphanumeric value

Found in MRCOC

Valid values for Source of co-occurrence:

MED02	MEDLINE (1997-2001)
MBD02	MEDLINE (1992-1996)
AIR93	AI/RHEUM
CCPSS99	Canonical Clinical Problem Statement System (CCPSS), 1999

SOUI Unique identifier of record in which the concept appears in source

Found in MRLO

An optional field.

SRL Source Restriction Level

Integer

Found in MRSO

The "Category of additional restrictions" of the vocabulary source of this string. See the UMLS License agreement for the meaning of each level, and for the restriction level for any source vocabulary.

Valid Values:

0 No additional restrictions.

1 LICENSEE is prohibited from translating the vocabulary source into another language or from producing other derivative works based on this single vocabulary source.

2 All category 1 restrictions AND

LICENSEE is prohibited from using the vocabulary source in operational applications that create records or information containing data from the vocabulary source. Use for data creation research or product development is allowed.

3 LICENSEE's right to use material from the source vocabulary is restricted to internal use at the LICENSEE's site(s) for research, product development, and statistical analysis only. Internal use includes use by employees, faculty, and students of a single institution at multiple sites. Notwithstanding the foregoing, use by students is limited to doing research under the direct supervision of faculty. Internal research, product development, and statistical analysis use expressly excludes: use of material from these copyrighted sources in routine patient data creation; incorporation of material from these copyrighted sources in any publicly accessible computer-based information system or public electronic bulletin board including the Internet; publishing or translating or creating derivative works from material from these copyrighted sources; selling, leasing, licensing, or otherwise making available material from these copyrighted works to any unauthorized party; and copying for any purpose except for back up or archival purposes.

LICENSEE may be required to display special copyright notices before displaying data from the vocabulary source. Applicable notices are included in the list of UMLS Metathesaurus Vocabulary sources, that is part of the "License Agreement for Use of UMLS Products".

STR String

Variable length string that includes alpha characters and may also include numerics, punctuation, or any character in the Metathesaurus character set

Found in MRCON and DELETED* files

A unique string or concept name that appears in one or more of the Metathesaurus source vocabularies. Any variation in upper-lower case is a separate string. The same string in different languages (e.g., English, Spanish) will have a different string identifier for each language.

STT String Type

Variable length string up to three characters

Found in MRCON

Valid values:

PF Preferred form of term OR

V Followed by one or more of the following types of variation, in this order:

- C Varies from the preferred term only in upper-lower case
- W Contains same words as the preferred form, disregarding order and punctuation
- S Singular of the preferred form
- P Plural of the preferred form
- O Other variant of the preferred form

An indication of whether the string is the preferred form of the term or a variant of that form. Variant tags (e.g., W,S,P) were assigned by computer algorithm.

NOTE: The Metathesaurus includes only those variants that are actually present in the Metathesaurus source vocabularies. The SPECIALIST Lexicon and Tools provide principled normalization functions which allow retrieval across lexical variation where specific variants do not exist in the source vocabularies. The algorithm used to identify lexical variants in the Metathesaurus is described in Section 4 and appears in the LEX directory.

MeSH includes a more complete set of lexical variants for its terms than other Metathesaurus sources. MeSH variants ("permuted terms", MSH/PM) are reviewed to exclude cases where the meaning is incorrect, although some inelegant forms exist.

STY Semantic Type

Found in MRSTY

Repeating element with the two subelements:

1. TUI Unique identifier for Semantic type
2. STY Semantic type

Valid Values for Unique identifiers

The identifiers in the UMLS Semantic Network.

Valid Values for Semantic Types:

The semantic types in the UMLS Semantic Network.

The basic category or categories to which the concept belongs, (e.g., Acquired Immunodeficiency Syndrome is a "Disease or Syndrome"), assigned by editors for all reviewed concepts. The Semantic type(s) are the link between the Metathesaurus and the UMLS Semantic Network. Application programs may be able to use the semantic type information to determine the general topics of interest in a particular user query.

SUI Unique Identifier for String

The letter S followed by seven digits

Found in MRCON, MRSAT, MRSO, MRCXT, MRLO, Word Indexes, MRXNW.ENG, MRXNS.ENG, AMBIG.SUI, and DELETED.SUI

The unique identifier for each string in the Metathesaurus. String identity is case sensitive - i.e., strings which differ only in case will receive different SUIs. This identifier remains the same across versions of the Metathesaurus. Each SUI is linked to a single LUI that represents the group of all strings that are lexical variants of each other. In cases where a single string can name multiple concepts, a SUI will be linked to multiple CUIs in the Metathesaurus.

SUPRES Suppressible Flag

One alphabetic character

Found in MRRANK

Flag indicating that this SAB and TTY will create a TS=s MRCON entry.

TS Term Status

One alpha character

Found in MRCON

Valid values:

P	Preferred Name
S	Synonym
s	Suppressible synonym

An indication of whether the term is the preferred name of the concept to which it is linked, or a synonym of the preferred name. There is a separate preferred name for each language.

Suppressible synonyms are less useful and possibly problematic for some applications, e.g. abbreviations, not fully specified or face-invalid names, which many users may wish to eliminate.

TTY Abbreviation for the type of name in a source vocabulary

Two characters

Found in MRSO and MRRANK

An abbreviation for the type of concept name in a Metathesaurus source vocabulary, for example PT for preferred terms. Allowed values are listed in Appendix B.4, Types of Names in a Vocabulary - the TTY and in the TTYDOC Bonus file available at:

http://umlsinfo.nlm.nih.gov

TUI Unique Identifier of Semantic Types

T followed by 3 numbers

Found in MRSTY

Valid values are found in the Semantic Network.

UN Meaning of frequency

Found in MRLO and is an optional field.

Valid value for meaning of frequency data:

*CITATIONS

Examples:

MBD02|12|*CITATIONS|

WD Word in Lowercase

Alphanumeric value

Found in Word Index files

XC Has Child

Found in MRCXT

A plus(+) sign indicates that the CUI2 for this row has children in this context. If this field is empty, the CUI2 does not have children in this context.

B.1.2 Attribute Descriptions

AM Ambiguous string indicator

The value format has two subelements, separated by a ":"

1. A single character: A

   An indicator that the case-insensitive string is ambiguous, i.e., has more than one meaning in the Metathesaurus. (See Section 2.2.2 and Figure 2 in this documentation). An ambiguous string has multiple entries in the Metathesaurus. These entries have the same CASE-INSENSITIVE string, but different concept identifiers (CUIs). Note that SUIs will differ where there are case differences.

2. The SUI of the corresponding disambiguating string (optional).

   The disambiguating strings are Metathesaurus "Preferred Names" (MTH/PN), created by Metathesaurus editors to help distinguish the meanings of identical strings.

Example: A:S0987654

AN MeSH Annotation

Variable length alphanumeric string that may contain punctuation.

An informative MeSH note written primarily for indexers or catalogers that may also be useful in explaining the use of a MeSH term to online searchers.

AQL Allowable Qualifier

A list of allowable qualifier abbreviations for MeSH main headings.

This information is also released in MRREL as AQ. In previous releases, this was only released as AQ.

Examples: AA, CL, CS, DF, DU, IM, I,P ME, PK

ATC Alternative Billing Codes Tree Code

Actual alternative medicine billing alphabetic code string. Treenumber for context in MRCXT.

CCF Canonical Clinical Problem Statement System (CCPSS) frequency

Numeric characters.

The number of times a CCPSS term appears in a patient record.

CCI ICD-9-CM code(s) clusters in a Clinical Classifications Software (CCS) category

Individual ICD-9-CM codes (or ranges of such codes) classified into CCS categories.

CFR Code of Federal Regulation Number

Numeric characters.

Sample of valid values:

862.3220
892.1610

CPA CPT Short Description

Variable length alphanumeric string with embedded punctuation.

The CPT abbreviated procedure description. (These are not unique, i.e., multiple long forms may have the same short form).

Examples:

ANESTH, TYMPANOTOMY

DIALYSIS, REPEATED EVAL.

CPF CPT Full Procedure

Variable length alphanumeric string with embedded punctuation.

The complete text of the CPT full procedure, in cases where the CPT2002 term in the "STR" field of MRCON has been trimmed from its original form. Trimming occurred in the modifiers and where the term exceeded 750 characters in length. Note that some full procedures may be several thousand characters long.

Example:

```
UNUSUAL PROCEDURAL SERVICES: When the service(s) provided is
greater
than that usually required for the listed procedure, it may be
identified by adding modifier '-22' to the usual procedure
number or by
use of the separate five digit modifier code 09922. A report
may also be
appropriate
```

CX Consider Also Note (MeSH only)

Variable length alphabetic string.

Other word roots or prefixes that should be consulted for concepts related to this MeSH concept, e.g., the value for "Heart" is "consider also terms at cardi- and myocardi-".

DA Date of entry

YYYYMMDD, e.g., 19920830

The date of entry of the concept into the Metathesaurus.

DC MeSH Descriptor class

A single numeric character. The type of MeSH term the concept name represents.

Valid Values:

1 = an Index Medicus descriptor

2 = a citation type descriptor (English Abstract)

3 = a check tag descriptor, e.g., HUMAN

4 = a geographic descriptor

DCSA Controlled Substance Act designation code

Numerical code.

Sample of Valid Values:

0
4

DDF Drug Doseform

Variable length alphabetic string.

Sample of Valid Values:

chewable tablet
inhaler

DDFA Drug Doseform Abbreviation

Variable length alphabetic string.

Sample of Valid Values:

SOLN
CAP

DHJC HCPCS J-code

The character J followed by numbers.

Multum clinical drugs linked to HCPCS J-codes where applicable.

Sample of Valid Values:

J7507
J7625

DPC Pregnancy Hazard Classification Code

Single alphabetic character.

Pregnancy Hazard Classification codes assigned to Multum clinical drugs.

Sample of Valid Values:

X
D

DQ Date Qualifier Established (MeSH only)

YYYYMMDD

The date the qualifier became available for indexing MEDLARS citations.

DRT Drug Route of Administration

Variable length alphabetic string.

Sample of Valid Values:

Injection (systemic)
Intramuscular
Intravenous

DRTA Drug Route of Administration Abbreviation

Variable length alphabetic string.

Sample of Valid Values:

INJ
TOP

DS MeSH Descriptor Sort Version

Variable length alphanumeric string.

The form needed for proper sequencing of the concept name, if the name could not be sequenced properly by the sort algorithms used in the MeSH publications.

Example:

MC Antithrombin III DS Antithrombin 03

DST Drug Strength

Variable length alphanumeric string with embedded punctuation.

Sample of Valid Values:

0.01%
0.02 MG
0.02 MG/ML

DX MeSH Date major descriptor established

YYYYMMDD

The first day of the Index Medicus publication month in which the descriptor (in any form) was available for searching as a major descriptor.

EC MeSH Entry combination

Variable length alphanumeric string with two subelements, separated by a colon (:)

1. Two character abbreviation for MeSH subheading
2. MeSH main heading

An invalid MeSH main heading/subheading combination that is a cross reference to a single MeSH main heading or a main heading/subheading combination that should be used in its place.

EV MeSH Entry term abbreviation

Variable length alphanumeric string.

A short form for a MeSH entry term or cross reference used primarily in MEDLINE record creation and maintenance.

EZ Enzyme Commission Number

A numeric string with embedded periods.

The International Union of Biochemists Enzyme Commission number for an enzyme concept. This number is an alias for the concept name and can be used to retrieve information about the concept from some databases. Enzyme commission numbers may also appear in the RN field, preceded by the designation EC.

FR MeSH Frequency

Numeric value

For MeSH supplementary concepts only, the number of times the chemical has been identified in articles indexed in MEDLINE as of January 2001.

FX MeSH MH Mapping

Maps a MeSH MH to a 'See Related' MH.

GAN Genbank Accession Number

Numeric value

HAB HCPCS abbreviation (short form)

Short descriptive text of procedure code (28 characters or less).

Standard alias:	HCPCS_SHRT_DESC_TXT
SAS alias:	SHRTDESC

Example: for the "long form" of:

D9110 | PALLIATIVE (EMERGENCY) TREATMENT OF DENTAL PAIN-MINOR PROCEDURES

There is a HAB of:

Tx dental pain minor proc

HAC HCPCS action code

A code denoting the change made to a procedure or modifier code within the HCPCS system. There are two subelements separated by "=". The first subelement is the HCPCS abbreviation of the value. The second subelement is the expanded value.

Standard alias: HCPCS_ACTN_CD

SAS alias: ACTN_CD

Valid Values:

- A = ADD PROCEDURE OR MODIFIER CODE
- B = CHANGE IN BOTH ADMINISTRATIVE DATA FIELD AND LONG DESCRIPTION OR PROCEDURE OR MODIFIER CODE
- C = CHANGE IN LONG DESCRIPTION OF PROCEDURE OR MODIFIER CODE
- D = DISCONTINUE PROCEDURE OR MODIFIER CODE
- F = CHANGE IN ADMINISTRATIVE DATA FIELD OF PROCEDURE OR MODIFIER CODE
- N = NO MAINTENANCE FOR THIS CODE
- P = PAYMENT CHANGE (MOG, PRICING INDICATOR CODES, ANESTHESIA BASE UNITS)
- R = RE-ACTIVATE DISCONTINUED/DELETED PROCEDURE
- S = CHANGE IN SHORT DESCRIPTION OF PROCEDURE CODE
- T = MISCELLANEOUS CHANGE (BETOS, TYPE OF SERVICE)

HAD HCPCS Action Effective Date

Effective data of action to a procedure or modifier code.

Standard alias: HCPCS_ACTN_EFCTV_DT

SAS alias: EFCTV_DT

Format: eight digits YYYYMMDD format

HAQ HCPCS Anesthesia Base Unit Quantity

The base unit represents the level of intensity for anesthesia procedure services that reflects all activities except time. These activities include usual pre-operative and post-operative visits, the administration of fluids and/or blood incident to anesthesia care, and monitoring procedures. (Note: The payment amount for anesthesia services is based on a calculation using base unit, time units, and the conversion factor).

Valid value: "0" (zero) or a three-digit number (i.e., 007)

HBT HCPCS Berenson-Eggers Type of Service Code

The Berenson-Eggers Type of Service (BETOS) for the procedure code based on generally agreed upon clinically meaningful groupings of procedures and services. There are two subelements separated by "=". The first subelement is the HCPCS abbreviation of the value. The second subelement is the expanded value.

Standard alias: HCPCS_BETOS_CD

SAS alias: BETOS

Title alias: BETOS_CD

Valid Values:

D1A	=	MED/SURG SUPPLIES
D1B	=	HOSPITAL BEDS
D1C	=	OXYGEN AND SUPPLIES
D1D	=	WHEELCHAIRS
D1E	=	OTHER DME
D1F	=	ORTHOTIC DEVICES
I1A	=	STANDARD IMAGING - CHEST
I1B	=	STANDARD IMAGING - MUSCULOSKELETAL
I1C	=	STANDARD IMAGING - BREAST
I1D	=	STANDARD IMAGING - CONTRAST G.I.
I1E	=	STANDARD IMAGING - NUCLEAR MEDICINE
I1F	=	STANDARD IMAGING - OTHER
I2A	=	ADVANCED IMAGING - CAT: HEAD

I2B = ADVANCED IMAGING - CAT: OTHER
I2C = ADVANCED IMAGING - MRI: BRAIN
I2D = ADVANCED IMAGING - OTHER
I3A = ECHOGRAPHY - EYE
I3B = ECHOGRAPHY - ABDOMEN/PELVIS
I3C = ECHOGRAPHY - HEART
I3D = ECHOGRAPHY - CAROTID ARTERIES
I3E = ECHOGRAPHY - PROSTATE - TRANSRECTAL
I3F = ECHOGRAPHY - OTHER
I4A = IMAGING/PROCEDURE - HEART- INC CARDIAC CATH
I4B = IMAGING/PROCEDURE - OTHER
M1A = OFFICE VISITS - NEW
M1B = OFFICE VISITS - ESTABLISHED
M2A = HOSPITAL VISIT - INITIAL
M2B = HOSPITAL VISIT - SUBSEQUENT
M2C = HOSPITAL VISIT - CRITICAL CARE
M3 = EMERGENCY ROOM VISIT
M4A = HOME VISIT
M4B = NURSING HOME VISIT
M5A = SPECIALIST - PATHOLOGY
M5B = SPECIALIST - PSYCHIATRY
M5C = SPECIALIST - OPHTHALMOLOGY
M5D = SPECIALIST - OTHER
M6 = CONSULTATIONS
O1A = AMBULANCE
O1B = CHIROPRACTIC
O1C = ENTERAL AND PARENTERAL
O1D = CHEMOTHERAPY
O1E = OTHER DRUGS
O1F = VISION - HEARING AND SPEECH SERVICES
O1G = INFLUENZA VACCINE
P0 = ANESTHESIA

P1A	=	MAJOR PROCEDURE - BREAST
P1B	=	MAJOR PROCEDURE - COLECTOMY
P1C	=	MAJOR PROCEDURE - CHOLECYSTECTOMY
P1D	=	MAJOR PROCEDURE - TURP
P1E	=	MAJOR PROCEDURE - HYSTERECTOMY
P1F	=	MAJOR PROCEDURE - EXPLOR/DECOMPR/EXCIS DISC
P1G	=	MAJOR PROCEDURE - OTHER
P2A	=	MAJOR PROCEDURE - CARDIOVASCULAR - CABG
P2B	=	MAJOR PROCEDURE - CARDIOVASCULAR - ANEURYSM REPAIR
P2C	=	MAJOR PROCEDURE - CARDIOVASCULAR - THROMBOENDARTERECTOMY
P2D	=	MAJOR PROCEDURE - CARDIOVASCULAR - CORONARY ANGIOPLASTY (PTCA)
P2E	=	MAJOR PROCEDURE - CARDIOVASCULAR - PACEMAKER INSERTION
P2F	=	MAJOR PROCEDURE - CARDIOVASCULAR - OTHER
P3A	=	MAJOR PROCEDURE - ORTHOPEDIC - HIP FRACTURE REPAIR
P3B	=	MAJOR PROCEDURE - ORTHOPEDIC - HIP REPLACEMENT
P3C	=	MAJOR PROCEDURE - ORTHOPEDIC - KNEE REPLACEMENT
P3D	=	MAJOR PROCEDURE - ORTHOPEDIC - OTHER
P4A	=	EYE PROCEDURES - CORNEAL TRANSPLANT
P4B	=	EYE PROCEDURES - CATARACT REM/LENS INS
P4C	=	EYE PROCEDURES - RETINAL DETACHMENT
P4D	=	EYE PROCEDURES - TREATMENT OF RETINAL LESIONS
P4E	=	EYE - OTHER
P5A	=	AMBULATORY PROCEDURES - SKIN
P5B	=	AMBULATORY PROCEDURES - MUSCULOSKELETAL
P5C	=	AMBULATORY PROCEDURES - GROIN HERNIA REPAIR
P5D	=	AMBULATORY PROCEDURES - LITHOTRIPSY
P5E	=	AMBULATORY PROCEDURES - OTHER
P6A	=	MINOR PROCEDURES - SKIN

P6B = MINOR PROCEDURES - MUSCULOSKELETAL
P6C = MINOR PROCEDURES - OTHER (MFS)
P6D = MINOR PROCEDURES - OTHER (NON MFS)
P7A = ONCOLOGY - RADIATION THERAPY
P7B = ONCOLOGY - OTHER
P8A = ENDOSCOPY - ARTHROSCOPY
P8B = ENDOSCOPY - UPPER G.I.
P8C = ENDOSCOPY - SIGMOIDOSCOPY
P8D = ENDOSCOPY - COLONOSCOPY
P8E = ENDOSCOPY - CYSTOSCOPY
P8F = ENDOSCOPY - BRONCHOSCOPY
P8G = ENDOSCOPY - LAPAROSCOPIC CHOLECYSTECTOMY
P8H = ENDOSCOPY - LARYNGOSCOPY
P8I = ENDOSCOPY - OTHER
P9A = DIALYSIS SERVICE (MFS)
P9B = DIALYSIS SERVICES (NOT MFS)
T1A = LAB TESTS - ROUTINE VENIPUNCTURE (NOT MFS)
T1B = LAB TESTS - AUTOMATED GENERAL PROFILES
T1C = LAB TESTS - URINALYSIS
T1D = LAB TESTS - BLOOD COUNTS
T1E = LAB TESTS -GLUCOSE
T1F = LAB TESTS - BACTERIAL CULTURES
T1G = LAB TEST - OTHER (MFS)
T1H = LAB TEST - OTHER (NON MFS)
T2A = OTHER TESTS - ELECTROCARDIOGRAMS
T2B = OTHER TESTS - CARDIOVASCULAR STRESS TESTS
T2C = OTHER TESTS - EKG MONITORING
T2D = OTHER TESTS - OTHER
Y1 = OTHER - MEDICARE FEE SCHEDULE
Y2 = OTHER - NON MEDICARE FEE SCHEDULE
Z1 = LOCAL CODES
Z2 = UNDEFINED CODES

HCC HCPCS Coverage Code

A code denoting Medicare coverage status. There are two subelements separated by "=". The first subelement is the HCPCS abbreviation of the value. The second subelement is the expanded value.

Standard alias: HCPCS_CVRG_CD

SAS alias: CVRG_CD

Valid Values:

C	=	CARRIER JUDGMENT
D	=	SPECIAL COVERAGE INSTRUCTIONS APPLY
G	=	NOT VALID FOR MEDICARE (90 DAY GRACE PERIOD)
I	=	NOT VALID FOR MEDICARE (NO GRACE PERIOD)
M	=	NON-COVERED BY MEDICARE
S	=	NON-COVERED BY MEDICARE STATUTE

HCD HCPCS Code Added Date

The year the HCPCS code was added to the HCFA Common Procedure Coding System.

Standard alias: HCPCS_CD_AD_DT

SAS alias: ADD_DT

Format: YY

HIR HCPCS Coverage Issues Manual Reference Section Number

Number identifying the Reference Section of the Coverage Issues Manual. Repeating element.

Standard alias: HCPCS_CIM_RFRNC_SECT_NUM

SAS alias: CIM

Example: 70-1; 70-2

HLC HCPCS Lab Certification Code

Code used to classify laboratory procedures according to the specialty certification categories listed by HCFA. Any generally certified laboratory (e.g.100) may perform any of the tests in its subgroups (e.g. 110, 120, etc.). There are two subelements separated by "=". The first subelement is the HCPCS abbreviation of the value. The second subelement is the expanded value.
Repeating element.

Standard alias: HCPCS_LAB_CRTFCTN_CD

SAS alias: LABCERT

Valid Values:

010	=	HISTOCOMPABILITY TESTING
100	=	MICROBIOLOGY
110	=	BACTERIOLOGY
115	=	MYCOBACTERIOLOGY
120	=	MYCOLOGY
130	=	PARASITOLOGY
140	=	VIROLOGY
150	=	OTHER MICROBIOLOGY
200	=	DIAGNOSTIC IMMUNOLOGY
210	=	SYPHILIS SEROLOGY
220	=	GENERAL IMMUNOLOGY
300	=	CHEMISTRY
310	=	ROUTINE CHEMISTRY
320	=	URINALYSIS
330	=	ENDOCRINOLOGY
340	=	TOXICOLOGY
350	=	OTHER CHEMISTRY
400	=	HEMATOLOGY

500	=	IMMUNOHEMATOLOGY
510	=	ABO GROUP & RH TYPE
520	=	ANTIBODY DETECTION (TRANSFUSION)
530	=	ANTIBODY DETECTION (NONTRANSFUSION)
540	=	ANTIBODY IDENTIFICATION
550	=	COMPATIBILITY TESTING
560	=	OTHER IMMUNOHEMATOLOGY
600	=	PATHOLOGY
610	=	HISTOPATHOLOGY
620	=	ORAL PATHOLOGY
630	=	CYTOLOGY
800	=	RADIOBIOASSAY
900	=	CLINICAL CYTOGENETICS

HM MeSH Heading Mapped To

The heading mapped to attribute in C-MeSH containing repeating (MH or MH/SH) elements.

Example: HM = PYRROLIDINONES
HM = *TARTRATES
HM = ESTRONE/* analogs & derivatives

HMP HCPCS Multiple Pricing Indicator Code

Code used to identify instances where a procedure could be priced**.** There are two subelements separated by "=". The first subelement is the HCPCS abbreviation of the value. The second subelement is the expanded value.

Standard alias: HCPCS_MLTPL_PRCNG_IND_CD

SAS alias: MULTCD

Valid Values:

9	=	NOT APPLICABLE AS HCPCS NOT PRICED SEPARATELY BY PART B (PRICING INDICATOR IS 00) OR VALUE IS NOT ESTABLISHED (PRICING INDICATOR IS '99')
A	=	NOT APPLICABLE AS HCPCS PRICED UNDER ONE METHODOLOGY
B	=	PROFESSIONAL COMPONENT OF HCPCS PRICED USING RVU'S, WHILE TECHNICAL COMPONENT AND GLOBAL SERVICE PRICED BY MEDICARE PART B CARRIERS
C	=	PHYSICIAN INTERPRETATION OF CLINICAL LAB SERVICE IS PRICED UNDER PHYSICIAN FEE SCHEDULE USING RVU'S, WHILE PRICING OF LAB SERVICE IS PAID UNDER CLINICAL LAB FEE SCHEDULE
D	=	SERVICE PERFORMED BY PHYSICIAN IS PRICED UNDER PHYSICIAN FEE SCHEDULE USING RVU'S, WHILE SERVICE PERFORMED BY CLINICAL PSYCHOLOGIST IS PRICED UNDER CLINICAL PSYCHOLOGIST FEE SCHEDULE
E	=	SERVICE PERFORMED BY PHYSICIAN IS PRICED UNDER PHYSICIAN FEE SCHEDULE USING RVU'S, SERVICE PERFORMED BY CLINICAL PSYCHOLOGIST IS PRICED UNDER CLINICAL PSYCHOLOGIST'S FEE SCHEDULE AND SERVICE PERFORMED BY CLINICAL SOCIAL WORKER IS PRICED UNDER CLINICAL SOCIAL WORKER FEE SCHEDULE
F	=	SERVICE PERFORMED BY PHYSICIAN IS PRICED UNDER PHYSICIAN FEE SCHEDULE BY CARRIERS. SERVICE PERFORMED BY CLINICAL PSYCHOLOGIST IS PRICED UNDER CLINICAL PSYCHOLOGIST'S FEE SCHEDULE AND SERVICE PERFORMED BY CLINICAL SOCIAL WORKER IS PRICED UNDER CLINICAL SOCIAL WORKER FEE SCHEDULE
G	=	CLINICAL LAB SERVICE PRICED UNDER REASONABLE CHARGE WHEN SERVICE IS SUBMITTED ON CLAIM WITH BLOOD PRODUCTS, WHILE SERVICE IS PRICED UNDER CLINICAL LAB FEE SCHEDULE WHEN THERE ARE NO BLOOD PRODUCTS ON CLAIM

HMR HCPCS Medicare Carriers Manual reference section number

Number identifying a section of the Medicare Carriers Manual
Repeating element.

Standard alias: HCPCS_MGM_RFRNC_SECT_NUM

SAS alias: MCM

Example: 2215

HN History Note

Variable length alphanumeric string with punctuation.

For MeSH history notes, the year when the current form of the MeSH term was established as a major and/or minor descriptor. The minor descriptor entry data appears in parentheses. Other brief information about history of the MeSH heading and its cross-references may follow the year data. MeSH terms that have been used continuously as major descriptors since 1963 do not have dates in this element.

The information in this field can be used to determine whether a MeSH term will appear in early MEDLINE backfiles and to find out how to search a concept in MEDLINE before its current term was introduced into MeSH.

For AOD history notes, a short text field describing the year the term was introduced or the history of the usage of the term.

Examples:

72 (68)

83

76 (75)

89; was /occurrence 196688

Changed descriptor 2000; through 1999 use "gestation."

HPD HCPCS ACD payment group effective date

The date the procedure is assigned to the ASC payment group.

Standard alias: HCPCS_ASC_PMT_GRP_EFCT_DT

SAS alias: ASCGRP

Eight digits YYYYMMDD

HPF HCPCS Full Code

Variable length alphanumeric string with embedded punctuation.

The complete text of the HCPCS string for a CPT modifier. The term in the "STR" field of MRCON has been trimmed from the original form (due to extraneous information at the end of the texts).

Example:

PROLONGED EVALUATION AND MANAGEMENT SERVICES: WHEN THE FACE-TO-FACE OR FLOOR/UNIT SERVICE (S) PROVIDED IS PROLONGED OR OTHERWISE GREATER THAN THAT USUALLY REQUIRED FOR THE HIGHEST LEVEL OF EVALUATION AND MANAGEMENT SERVICE WITHIN A GIVEN CATEGORY, IT MAY BE IDENTIFIED BY ADDING MODIFIER -21 TO THE EVALUATION AND MANAGEMENT CODE NUMBER OR BY USE OF THE SEPARATE FIVE DIGIT MODIFIER CODE 09921. A REPORT MAY ALSO BE APPROPRIATE.

HPG HCPCS ASC payment group code

Standard alias: HCPCS_ASC_PMT_GRP_CO

SAS alias: ASCIND

Edit rules: Range 01 to 09

The code which represents the dollar amount of the facility charge payable by Medicare for the procedure. Payment group rates, which are updated annually (most recently on October 1, 1997), are as follow:

Group 1 - $314
Group 2 - $422
Group 3 - $482
Group 4 - $595
Group 5 - $678
Group 6 - $789 ($639+$150)*
Group 7 - $941

Group 8 - $928 ($778+$150)*

*The $150 payment allowance in Group 6 and 8 is for intraocular lenses.

HPI HCPCS Pricing Indicator Code

Code used to identify the appropriate methodology for developing unique pricing amounts under Part B. A procedure may have one to four pricing codes. The identification is not all inclusive, as it excludes those prices where methodology is also identified by type of provider/supplier (i.e. nurse practitioner, physician assistant, clinical psychologist, clinical social worker, etc.). There are two subelements separated by "=". The first subelement is the HCPCS abbreviation of the value. The second subelement is the expanded value.
Repeating element.

Standard alias: HCPCS_PRCNG_IND_CD

SAS alias: PRCNGCD

Valid Values:

00	=	SERVICE NOT SEPARATELY PRICED BY PART B (E.G., SERVICES NOT COVERED, BUNDLED, USED BY PART A ONLY, ETC.)
11	=	PRICE ESTABLISHED USING NATIONAL RVU'S
12	=	PRICE ESTABLISHED USING NATIONAL ANESTHESIA BASE UNITS
13	=	PRICE ESTABLISHED BY CARRIERS (E.G., NOT OTHERWISE CLASSIFIED, INDIVIDUAL DETERMINATION, CARRIER DISCRETION)
21	=	PRICE SUBJECT TO NATIONAL LIMITATION AMOUNT
22	=	PRICE ESTABLISHED BY CARRIERS (E.G., GAP-FILLS, CARRIERS ESTABLISHED PANELS)
31	=	FREQUENTLY SERVICED DME (PRICE SUBJECT TO FLOORS AND CEILINGS
32	=	INEXPENSIVE AND ROUTINELY PURCHASED DME (PRICE SUBJECT TO FLOORS AND CEILINGS)
33	=	OXYGEN AND OXYGEN EQUIPMENT (PRICE SUBJECT TO FLOORS AND CEILINGS)

34	=	DME SUPPLIES (PRICE SUBJECT TO FLOORS AND CEILINGS)
35	=	SURGICAL DRESSINGS (PRICE SUBJECT TO FLOORS AND CEILINGS)
36	=	CAPPED RENTAL DME (PRICE SUBJECT TO FLOORS AND CEILINGS)
37	=	OSTOMY, TRACHEOSTOMY AND UROLOGICAL SUPPLIES (PRICE SUBJECT TO FLOORS AND CEILINGS)
38	=	ORTHOTICS, PROSTHETICS, PROSTHETIC DEVICES & VISION SERVICES (PRICE SUBJECT TO FLOORS AND CEILINGS)
45	=	CUSTOMIZED DME ITEMS
46	=	CARRIER PRICED (E.G., NOT OTHERWISE CLASSIFIED, INDIVIDUAL DETERMINATION, CARRIER DISCRETION, GAP-FILLED AMOUNTS)
51	=	DRUGS
52	=	REASONABLE CHARGE
53	=	STATUTE
54	=	VACCINATIONS
55	=	PRICED BY CARRIERS UNDER CLINICAL PSYCHOLOGIST FEE SCHEDULE
56	=	PRICED BY CARRIERS UNDER CLINICAL SOCIAL WORKER FEE SCHEDULE
57	=	OTHER CARRIER PRICED
99	=	VALUE NOT ESTABLISHED

HPN HCPCS processing note number

The number identifying the processing note contained in Appendix A of the HCPCS Manual.

Standard alias: HCPCS_PRCSG_NOTE_NUM

SAS alias: PROCNOTE

example: 0041

HSN HCPCS Statute Number

Number identifying statute reference for coverage or noncoverage of procedure or service.

Standard alias: HCPCS_STATUTE_NUM

SAS alias: STATUTE

example: 1862 A(12)

HTD HCPCS Termination Date

Last date for which a procedure or code may be used by Medicare Providers.

Standard alias: HCPCS_TRMNTN_DT

SAS alias: TERM_DT

Eight Digits YYYYMMDD

HTS HCPCS Type of Service Code

The carrier assigned HCFA Type of Service which describes the particular kind(s) of service represented by the procedure code. There are two subelements separated by "=". The first subelement is the HCPCS abbreviation of the value. The second subelement is the expanded value.
Repeating element.

Standard alias: HCPCS_TYPE_SRVC_CD

SAS alias: TYPESRVC

Valid Values:

0	=	WHOLE BLOOD OR PACKED RED CELLS
1	=	MEDICAL CARE

2	=	SURGERY
3	=	CONSULTATION
4	=	DIAGNOSTIC RADIOLOGY
5	=	DIAGNOSTIC LABORATORY
6	=	THERAPEUTIC RADIOLOGY
7	=	ANESTHESIA
8	=	ASSISTANT AT SURGERY
9	=	OTHER MEDICAL ITEMS OR SERVICES
A	=	USED DURABLE MEDICAL EQUIPMENT (DME)
B	=	HIGH RISK SCREENING MAMMOGRAPHY
C	=	LOW RISK SCREENING MAMMOGRAPHY
D	=	AMBULANCE (EFF 04/95)
E	=	ENTERAL/PARENTERAL NUTRIENTS/SUPPLIES (EFF 04/95)
F	=	AMBULATORY SURGICAL CENTER (FACILITY USAGE FOR SURGICAL SERVICES)
G	=	IMMUNOSUPPRESSIVE DRUGS
H	=	HOSPICE SERVICE (DISCONTINUED 01/95)
I	=	PURCHASE OF DME (INSTALLMENT BASIS) (DISCONTINUED 04/95)
J	=	DIABETIC SHOES (EFF 04/95)
K	=	HEARING ITEMS AND SERVICES (EFF 04/95)
L	=	ESRD SUPPLIES (EFF 04/95) (RENAL SUPPLIER IN THE HOME BEFORE 04/95)
M	=	MONTHLY CAPATAION FOR DIALYSIS
N	=	KIDNEY DONOR
P	=	LUMP SUM PURCHASE OF DME, PROSTHETICS, ORTHOTICS
Q	=	VISION ITEMS OR SERVICES
R	=	RENTAL OF DME
S	=	SURGICAL DRESSINGS OR OTHER MEDICAL SUPPLIES (EFF 04/95)
T	=	PSYCHOLOGICAL THERAPY
U	=	OCCUPATIONAL THERAPY
V	=	PNEUMOCOCCAL/FLU/HEPATITIS B VACCINE (EFF 04/95)

(HEPATITIS ONLY BEFORE 04/95)

W = PHYSICAL THERAPY

Y = SECOND OPINION ON ELECTIVE SURGERY

Z = THIRD OPINION ON ELECTIVE SURGERY

HXR HCPCS Cross reference code

An explicit reference crosswalking a deleted code or a code that is not valid for Medicare to a valid current code (or range of codes).
Repeating element.

Standard alias: HCPCS_XREF_CD

SAS alias: XREF

IAA ICD10AM Abbreviated Descriptor

Variable length alphanumeric string with embedded punctuation.

Example:

C0024117|L0783000|S0837811|J44.9|IAA|ICD10AM|COPD unspecified|

IAC ICD10AM Australian Code

One character numeric string.

Contains a flag "1" for codes that have been exclusively developed in Australia.

IAD ICD10AM Effective from:

Variable length numeric string with embedded punctuation.

Indicates the date that the code is effective from.

Example:

1/7/98

IAH ICD10AM Age Edit, higher limit

Three character numeric string.

Minimum age is expressed as a three digit field.

Example:

Ages are represented as:

000-006: 0-6 days
011-013: 1-3 weeks
101-111: 1-11 months
201-299: 1-99 years
300- : 100+ years

IAL ICD10AM Age Edit, lower limit

Three character numeric string.

Minimum age is expressed as a three digit field.

Example:

Ages are represented as:

000-006: 0-6 days
011-013: 1-3 weeks
101-111: 1-11 months
201-299: 1-99 years
300- : 100+ years

IAN ICD10AM Annotation Note

One character string.

```
+ (Dagger) denotes a code describing the etiology or underlying cause
of the disease

* (asterisk) denotes a code describing the manifestation of a disease

The annotations of * (asterisk) and + (dagger) indicate those disease
codes in
```

```
ICD-10-AM which need another code (either asterisk or dagger) to
complete the
diagnostic statement.  Dagger codes are always sequenced before the
accompanying
asterisk code.
```

Example:

> "anterior spinal and vertebral artery compression syndromes" is coded as M47.0+ (sequenced first) followed by G99.2*.

IAR ICD10AM Relationship Note

Variable length alphanumeric string with embedded punctuation.

A code embedded in the term, may have a * or +. Any embedded codes that could be mapped 1-1 were included in MRREL. All also end up in MRSAT as IAR.

Example:

> C95.0

IAS ICD10AM Sex Edit

One character numeric string.

An edit flag to indicate whether the code is valid for a particular sex. Sex flags are 1 (male) or 2 (female).

IAT ICD10AM Sex Edit Type

One character numeric string.

All sex flagged codes are followed by a "sex edit type" flag. The sex edit type flags are 1 (fatal) or 2 (warning).

IAY ICD10AM Age Edit Type

One character numeric string.

All age flagged codes are followed by an age edit type flag. The age edit type flags are 1 (fatal) or 2 (warning).

ICA ICD Additional Codes Note

Variable length alphanumeric string with embedded punctuation.

An ICD-9-CM instruction to signal the coder that an additional code should be used if the information is available to provide a more complete picture of that diagnoses or procedure.

Example:

Use additional code to identify any associated silicosis(502)

ICC ICD Codes Also Note

Variable length alphanumeric string with embedded punctuation. Brackets {} may be used to indicate hierarchical information.

An ICD-9-CM instruction to signal the coder to code additional information.

Example:

Code also any synchronous debridement of brain

ICE ICD Entry Term

Variable length alphanumeric string. Brackets {} may be used to indicate hierarchical information.

An ICD-9-CM entry term.

Example:

Diarrhea:{dysenteric; epidemic}; Infectious diarrheal disease.

ICF ICD Fifth-digit Code Note

Variable length alphanumeric string with embedded punctuation. Brackets {} may be used to indicate hierarchical information.

An ICD-9-CM instruction to coders indicating use of a fifth-digit code.

Example:

> The following fifth-digit subclassification is for use with category 204: {0 without mention of remission; 1 in remission}

ICN ICD Note

Variable length alphanumeric string with embedded punctuation.

An ICD-9-CM instruction providing additional coding information.

Example:

> Requires fifth digit. See beginning of Section 010-018 for codes and definitions.

ICS ICD Short Form

Variable length alphanumeric string with embedded punctuation.

An ICD-9-CM 25-character version of the code.

II MeSH Indexing Information

One or more MeSH main headings or Main heading/subheading combinations.

For MeSH chemical terms (Term Type=NM), MeSH headings that may be relevant to articles that are also assigned the NM term. The information in this field helps to define the NM and suggests other related search terms.

INC ICD10AM Neoplasm code

Variable length alphanumeric string with embedded punctuation.

Example: C34.-

IPX Files linking ICPC to ICD10

A repeatable field

+ indicates that the ICD10 code is broader than the ICPC code.

- indicates that the ICD10 code is narrower than the ICPC code.

Example:

> The ICPC term: L81|Other injury has one IPX value of : -S47
> (S47=Crushing injury of shoulder and upper arm)

Some of the values have an "R" after the decimal point.

Example: -S93.R

IXR ICPC2E to ICD10 Cross Reference

Variable length alphanumeric string with embedded punctuation.

Example:

> ICD10: Q60, Q61, Q62, Q63, Q64

JXR Cross reference from a Jablonski term to an OMIM code.

Variable length alphanumeric string with embedded punctuation.

Jablonski terms that are linked to OMIM codes.

Sample of Valid Values:

> OMIM: 119600
> OMIM: 104760.0009

LAC ASTM Code

A numeric string of up to eight characters that includes an embedded period.

The ASTM E1238-94 code for a test in a LOINC name.

LAL LOINC Answerlist

Variable length alphanumeric string with embedded semi-colons separating different values in the list.

Examples:

GIVEN;PARTIALLY GIVEN;DISCARDED
AUTOLOGOUS;DIRECTED DONOR;ANONYMOUS DONOR

The list of answers for results that are reportable from a multiple choice list, e.g., the answers for the term DISPOSITION OF BLOOD PACK are GIVEN;PARTIALLY GIVEN;DISCARDED.

LCA LOINC ACSSYM field

Variable length alphanumeric string with embedded punctuation.

LOINC chemical name synonyms, alternative names and chemical formulae from the Chemical Abstract Society.

LCB LOINC Chemical base name

Variable length alphanumeric string with embedded punctuation.

Chemical base name from the Chemical Abstract Society.

LCC LOINC CDC Code

Four numeric characters.

The code from the Centers for Disease Control Complexity file that maps a laboratory test to the instruments used to perform the test. The code is at the analyte level, not at the test instrument level.

LCI LOINC Molecular structure ID

The molecular structure ID, usually a Chemical Abstract Society number.

LCL LOINC Class

An alphabetic string.

Sample of valid values:

ABXBACT	Antibiotic susceptibility
ALLERGY	Response to antigens
BC	Cell counts (blood, CSF, pleuritic fluid)
BDYCRC	Body circumference

BDYHGT	Body height
BDYSURF	Body surface area
BDYTMP	Body temperature
BDYWGT	Body weight
BLDBK	Blood bank
BP	Blood pressure
BP.CENT	Blood pressure - central
BP.PSTN	Blood pressure - positional
BP.TIMED	Blood pressure - timed
BP.VENOUS	Blood pressure - venous
CELLMARK	Cell surface models
CHAL	Challenge tests
CHALSKIN	Skin challenge tests
CHEM	Chemistry
CLIN	Clinical NEC
COAG	Coagulation study
CYTO	Cytology
DRUG	Drug levels
DRUGDOSE	Drug dose (for transmitting doses for pharmacokinetics)
ED	Emergency department
EKG	Electrocardiogram
EKG.IMP	Electrocardiogram impression
EKG.MEAS	Electrocardiogram measures
EYE	Eye
FERT	Fertility
FUNCTION	Functional status (e.g. Glasgow)
H&P	History and physical
HEM	Hematology (excluding coagulation and differential count)
HEMODYN	Hemodynamics
HLA	HLA tissue typing antigens
HRTRATE	Heart rate
IO	Input/Output

MICRO	Microbiology
NEONAT	Neonatal measures
OB.US	Obstetric ultrasound
OBGYN	Obstetrics/gynecology
PATH	Pathology
RESP	Respiration
SERO	Serology (antibodies and most antigens except blood bank and infectious agents)
SKNFLD	Skinfold measurements
SURGPATH	Surgical pathology
TOX	Toxicology
UA	Urinalysis
VOLUME	Volume of specimens

An arbitrary classification of terms in LOINC designed to assist LOINC development and to group related observations together.

LCN LOINC Classtype

Valid Values:

1 = Laboratory class

2 = Clinical class

LCR LOINC Reason for Change

Variable length alphanumeric string with embedded punctuation.

Example:

CHANGED NAME -- WAS BASOPHILS/100 WBCS; CHANGED PROPERTY -- WASPCT

A brief explanation of the change made to a LOINC term.

LCS LOINC Depreciated or superseded status

Three alphabetic characters: DEL

An indicator that a LOINC term is no longer to be used. The term that should now be used will appear in the LMP element.

LCT LOINC Change Type Code

Three alphabetic characters. The type of change made to a LOINC term.

Valid Values:

ADD = Add

DEL = Delete

MAJ = Change to other subfield(s) [#2-#6] of fully-specified name

MIN = Change to any other LOINC field or attribute

NAM = Change to analyte/component, subfield #1 of fully-specified name

---

LDE LOINC DEEDS_CD

Variable length alphanumeric string with embedded punctuation.

Data Elements for Emergency Department Systems Codes (CDC). This field contains the DEEDS code value which maps to the LOINC code in question.

Example:

PE02000
4.20

LEA LOINC Example Answers

For some tests and measurements, LOINC has supplied examples of valid answers. These values differ from those in the ANSWERLIST field because that details possible values for nominal scale terms.

Sample of valid values:

1:64
negative @ 1:16

LEC LOINC Analyte Code

The letter A followed by five numbers.

The EUCLIDES code for the analyte which is the first subpart of the first part of a LOINC name.

LFO LOINC Formula

Regression equation details for many OB.US calculated terms.

Example:
DIAMETER.BIPARIETAL [cm] = -3.08 + 0.41 * (gestational age [wk]) - 0.000

LFR LOINC French name for term

French name for LOINC term supplied by Centre Suisse de Controle de Qualite. This field contains extended characters and will not transfer to 7-bit systems.

LGC GPI Code, GPI Code Total

For drugs, this field contains a map to the Medispan GPI codes, a hierarchical system of classifying pharmaceutical products. For a few products, a simple one-to-one mapping with a GIP code was not possible. In these cases, all applicable GPI codes are contained in this field, separated by semicolons.

Example:

7320001010

LGR German Name for LOINC Term

Supplied by Centre Suisse de Controle de Qualite. This field contains extended characters and will not transfer correctly to 7-bit systems

Example:

Ex--Keime; ; aerobische Kultur
Ex--Keime; Routina Bakterienkultur

LIC IUPAC Code

The letters QU followed by five numbers.

The IUPAC code for the component, kind of property, and system in a LOINC name. Note that most IUPAC codes assume that the component is measured in substance concentration, e.g., moles. The IUPAC code for substance concentration is applied to mass concentration in LOINC, because IUPAC has no code for the mass concentration variant, which is more commonly used in the U.S.

LIR Italian Name for LOINC Term

Supplied by Centre Suisse de Controle de Qualite. This field contains extended characters and will not transfer correctly to 7-bit systems

Example:

isolato--amossicillina
isolato--amossicillina+clavulanato
isolato--amfotericina B

LIU IUPAC Analyte code

This field contains the chemical abstract service number or the enzyme nomenclature number for the chemical components for chemicals and/or enzymes. These were also contributed by IUPAC.

Example:

CAS2022-85-7

LLR LOINC Date Last Changed

YYYYMMDD

The date the LOINC term was last changed.

LMC LOINC Metpath Code

Four numeric characters.

The code that represents the LOINC name at MetPath laboratories.

LMM Molecular weights

This field contains the molecular weights of many chemical moieties when they are provided. Those added in this release were kindly contributed by IUPAC.

Example: 204.31

LMP LOINC Map to Code

A numeric string with an embedded hyphen.

The LOINC code of the term that has superseded a term with a LCS value of DEL.

LMT LOINC MULTUM_CD

Maps to Multum Inc. database of codes for drugs.

Example:
d00226

LNC LOINC NAACCR_ID

Numeric value

LOINC terms mapped to North American Association of Central Cancer Registries Identification

---

LNE LOINC CODE_TABLE

The characters CR followed by numbers.

Examples from Cancer Registry

CR1930
CR650

LPL LOINC Panel Elements

Alphanumeric text, repeating field, separated by ';'.

List of individual tests that comprise a panel.

Example: "CHLORIDE;POTASSIUM;SODIUM"

LQS Survey Question Source

Exact name of the survey instrument and the item/question number.

Example:

> HOME HEALTH CARE CLASSIFICATION O38.1
> HOME HEALTH CARE CLASSIFICATION A01.0

LQT Survey Question Text

Verbatim question from the survey instrument.

Example:

> LIVING WITH HIV/AIDS -I MAY CHOOSE TO SELF-DELIVER (SUICIDE) NEAR THE END
> LIVING WITH HIV/AIDS -I HAVE A MEANINGFUL LIFE
> LIVING WITH HIV/AIDS -MY MEDICINES ARE TAKING OVER MY LIFE

LRF Reference

Contains references to medical literature, product announcements, or other written sources of information on the test or measurement described by the LOINC record.

LSC LOINC SNOMED Code

An alphabetic character followed by numeric characters and an embedded hyphen.

The code for a SNOMED International laboratory procedure name that is related to (usually broader than) the LOINC term.

LSP LOINC Species code

2 alphabetic characters.

Codes detailing which non-human species a term applies to. If there is no code, human is assumed.

LSR Root of a set of LOINC codes

Currently used for claims attachments.

Yes in this field signifies that this record is the root of a set of LOINC codes.

LUN LOINC Typical Units

Alphanumeric string with embedded slash.

Example: KIU/L

The typical units in which the observation is recorded.

LUR Units required when used as OBX segment

Y/N field that indicates that units are required when this LOINC is included as an OBX segment in a HIPAA attachment

---

MDA MeSH date of entry

YYYYMMDD

The date the term was added to the MeSH file, which is prior to the date the term became available for indexing and searching MEDLARS citations. Terms that have been part of MeSH for many years may have no value in this element.

MEA In NOC, the "measurement scale" used for a particular outcome.

Example:

9: None - Limited - Moderate - Substantial - Extensive

MED#### (MEDLINE Backfile Postings 1966 through 2002 (MeSH only)

'MED' followed by a numeric value followed by another numeric value preceded by an asterisk (*); one of the two values may be absent, indicating no occurrences.

Data elements which give the number of times a MeSH term has been used in MEDLINE, by year on the NLM system as of Fall 2001. For MeSH main headings, each element has two values: the total occurrences and the occurrences in which the term was designated a principal concept, preceded by an asterisk. For subheadings, only the total postings are given. These attributes have a SAB of MED02.

This is the only source of information in the Metathesaurus on the frequency of use of MeSH headings in MEDLINE prior to 1985 and on the total frequency of use of MeSH headings in MEDLINE from 1966 to Fall 2001.

MMR MeSH revision date

YYYYMMDD

The date of the last major revision to the term's MeSH record.

MN MeSH Tree Number

Repeating element. Variable length alphanumeric string with embedded periods.

The hierarchical number for the concept in the MeSH tree structures. This number also appears in the HCD subelement of the REL and CXT elements.

MR Major revision date

YYYYMMDD

The date the Metathesaurus entry for the concept underwent any revision in content.

MSC Minimal Standard (Terminology) Class

A variable length character string with embedded punctuation.

Example:

Reason.Term
Site.Term
Procedure.Attribute.Value

MSP SPN Medical Specialty Panel

2 alphabetic characters.

Medical Specialty Panel (responsible for reviewing the product).

MUI MeSH Unique Identifier

8 char Text field with leading 'M'.

The MUI is a MeSH Unique Identifier assigned by NLM. Each concept in MeSH is given a MUI.

Example: M0001333

NA Neuronames Abbreviation

One or more alphabetic characters

A short abbreviation for a concept name in the Neuronames thesaurus.

NAF NIC atoms

The complete text of the NIC full intervention, in cases where the NIC99 term in the "STR" field of MRCON has been trimmed from its original form (due to length or to extraneous information at the end of the text).

SA (the trimmed version) and AL (the full version) always stay together.

Examples:

44301016 | NIC99 | NIC99/AL | U001016 | Assist with fetal diagnostic procedures (e.g., amniocentesis, chorionic villus sampling, percutaneous umbilical blood sampling, and Doppler blood flow studies)

443086932 | NIC99 | NIC99/SA |U001016 | Assist with fetal diagnostic procedures

44306769|NIC99|NIC99/AL|U006769|Pro vide community resource information to elder patients and their caretakers,

including addresses and phone numbers of agencies that provide senior service assistance; health, home health, residential, respite, and emergency care; housing assistance; transportation; substance abuse treatment; sliding-fee counseling services; food pantries and Meals on Wheels; clothing distribution centers; human services; and hot lines

44308999|NIC99|NIC99/SA|U006769|Provide community resource information to elder patients and their caretakers

---

NAT Neuronames Anatomy Type

A single character.

Valid Values:

S superficial
V volumetric

An indication of the type of anatomy represented by a Neuronames concept name.

NFI National formulary indicator

Alphabetic characters.

Valid Values:

"YES" or "NO" indicating whether a drug is in the VA's National Formulary

---

NH NonHuman Flag

A single character: Y.

An indication that the concept does not apply to human beings, used only when the concept's Semantic type(s) could imply the contrary. For example, the concept BEAK and CLAW are assigned the Semantic type" Body Part, Organ, or Organ Component", but do not apply to human beings. Concepts do not carry the non human flag if their

semantic types are obviously nonhuman (e.g., "Medical Device", "Hazardous or Poisonous Substance"). This element can indicate when special search restrictions are appropriate or that a concept may be inappropriate to the user's search.

STY's of concepts eligible to receive the non-human flag are:

1. Anything in the Anatomical Structure Tree, e.g., horn
2. Anything in the Behavior Group, e.g., rutting
3. Anything in the Natural Phenomenon or Process Tree, e.g., glanders

---

NSR Neuronames Species Restriction

A single character.

Valid Values:

H human only
M macaque only

An indication that a Neuronames concept applies only to humans or only to macaques. Most Neuronames concepts apply to both and have no species restriction.

NST Normalized Drug Strength Alphanumeric string

Alphanumeric characters with punctuation.

Example:

PWDR,RENST-ORAL
27 MG/100ML

OL MeSH Online Note

Variable length alphanumeric string with punctuation.

Information helpful to online searchers of MEDLINE, especially when the history of a term or cross-reference has implications for online searching. This is a potential source of useful information for rules for search interface programs.

PA MeSH Pharmacologic Action

The pharmacologic action of MeSH main headings (MH) for drugs and supplementary concept names (NM). The information in this element is also represented by an "isa" relationship between the MH or NM concept and the MeSH concept name for the class of drugs with a particular pharmacologic action.

PCD PDQ Closest Related Clinical Diagnoses

Used for disease terms only.

The closest related clinical diagnosis is provided for disease terms to permit the identification of the closest diagnosis for which an entire treatment statement can be found in PDQ.

Example:
AIDS-related lymphoma

PCL Pharmacy Practice Activity Classification (PPAC) Category

All terms are assigned to one of five categories, which connote their hierarchy.

Valid Values:

Activity
Class
Domain
Step
Task

PDA PDQ Short Name or Abbreviation

An alphanumeric string with embedded blanks.

An abbreviated name for a PDQ concept.

PDC SPN Product Device Class

1 numeric character.

Product Device Class (level of CDRH regulation: class 1, 2, or 3).

PI MeSH Previous Indexing

Repeating element.

MeSH heading or heading/subheading combination(s) followed by a date range in parentheses (YYYY).

The MeSH headings or main heading/subheading combination that may be used to search for MEDLINE citations indexed before the concept was introduced into MeSH. Each entry is followed by a year or range of years as a guide to the period when the term in the PI field should be searched. The appropriate Boolean logic for the PI terms is not provided. The searcher must make the decision whether to AND or OR the terms supplied when multiple terms are present.

PLR Pharmacy Practice Activity Classification (PPAC) Last Revision Date

Format: M/DD/YY time

Example:
1/15/98 0:00:00

PM Public MeSH note

Variable length alphanumeric.

Combines key information from the HN and PI elements in a format that is printed in the MeSH publications.

PRC Product Third Party Review Code

1 alphabetic character.

Third party review code from SPN.

PTR SPN Product Tier

1 alphabetic or numeric character.

Product Tier (level of CDRH triage: 1, 2, 3, or E{xempt}).

PXC PDQ Protocol Exclusion Criteria

Terms with type "exclusion criteria," which may be indexed on protocol records to identify conditions that exclude a patient from eligibility.

Example:

active peptic ulcer disease/GI bleeding

PYR PsychInfo year designation

4-digit year

Example: 1973

QA MeSH Topical Qualifier Abbreviation

Two alphabetic characters.

For MeSH subheadings (Term Type=TQ), an abbreviation that may be used in place of the full text of the abbreviation in searching on NLM's system and possibly on other systems offering NLM data.

QE MeSH Qualifier Entry Version

Variable length alphabetic string

A short form for a MeSH qualifier.

QS MeSH Qualifier Sort Version

Alphabetic string.

The form of the subheading needed for proper alphabetic sequencing when the subheading cannot be sequenced properly by the sort algorithms used in the MeSH publications.

REF A list of bibliographic citations related to a given nursing intervention or nursing outcome.

Example:

Glick, O.J. (1992), Interventions related to activity and movement. In G.M. Bulechek & J.C. McCloskey (Eds.), Symposium on Nursing Interventions. Nursing Clinics of North America, 27 (2), 541-568.; Lewis,

C. B. (1989). Improving mobility in older persons. Rockville, MD: Aspen.;
Sheahan, S. (1982). Assessment of low back pain. Nurse Practitioner, 7 15-23.;
Sweezey, S. (1988), Low back pain. Geriatrics, 43(2), 39-44.

RID Read Codes Term_id

An identifier assigned to a Read term, used in referring to the term in the Read file structure, and may be used in clinical information systems.

RN Registry Number

A series of numbers and hyphens (any leading zeros in an RN are dropped) or a series of numbers and periods, preceded by EC

The Chemical Abstracts Service Registry number for a chemical or the Enzyme Commission number for an enzyme. (See also EZ.) This number is an alias for the concept name and can be used to retrieve information about the concept from a number of chemical and toxicological databases.

RR Related Registry Number

Repeating element; with two subelements.

Registry number (Relationship to concept in parentheses).

Example:

80551640 (sulfate)

The Chemical Abstracts Registry numbers for salts, optical isomers, or isotope-labeled versions of the concept followed by the relationship of this RR to the RN (in parentheses.) Applies to chemicals only. These numbers can be used as links to information in a number of chemical and toxicological databases.

RXO Drug Description from NDF

Single Character.

R for Rx and O for OTC

SB SNOMED International subset indicator

A one or two character code including letters and/or an asterisk.

Valid Values:

*	can code using two T codes or G code for laterality B Bethesda system (Morphology)
IC	ICDO (Oncology) related
N	Nursing
N*	Nursing, provisional
U	Ultrastructure (Morphology)
V	Veterinary
V*	Veterinary AND can code using two T codes or G code for laterality (Topography)

A tag indicating that a SNOMED International term belongs to one of a number of special categories.

SHF SNOMED Hospital Formulary Code.

A numeric string that includes colons and periods.

Example:

84:24:12

The American Hospital Formulary Code for a chemical contained in SNOMED International.

SIC SNOMED ICD9CM Reference

Repeating element; a numeric string which may have embedded periods.

The ICD9CM code or codes listed as relevant to the meaning of the concept in SNOMED International. The information contained in this element is also represented as SNMI-specific "mapped_to" relationships between Metathesaurus concepts in the MRREL file.

SMX SNOMED Multiaxial coding

An alphanumeric string that includes hyphens, parentheses, and sometimes ellipses.

Additional SNOMED International codes or truncated codes that cover aspects of the SNOMED International concept to which they are attached.

SOS Scope Statement

Two subelements:

1. Abbreviation of the source of the statement
2. Variable length string with alpha character, punctuation, and in some cases diacritics

Details about the scope of a concept in a particular source vocabulary. The Metathesaurus contains scope statements from a number of source vocabularies.

SRC MeSH Literature source of chemical name

Variable length alphanumeric string.

A citation to an article in a journal indexed for MEDLINE in which the chemical has been identified.

Note: Not to be confused with source abbreviation of SRC.

ST Concept Attributes Status

One alpha character

Valid Values:

R Reviewed

U Unreviewed

The review status of the concept attributes. If human review of the concept attributes has been completed, the status is R. If human review has not been completed, the status is U.

All concepts in the current release are reviewed, so there are no Unreviewed cases.

Note: In prior releases of the Metathesaurus, some concepts (primarily MeSH Supplementary Concepts) were "Unreviewed" because they were not reviewed as concepts for the Metathesaurus. All have been reviewed as concepts for the current release.

Similarly, some interconcept relationships derived from the Supplementary Concepts were not reviewed again for the Metathesaurus; these relationships carried the "Machine-generated and unverified indicator" (MG, value G). All relationships have been reviewed for the current release.

SWP Swiss Protein Number

Numeric value

TH MeSH Thesaurus ID

Repeating element. Alphanumeric string.

Identifies thesauri other than MeSH in which the MeSH heading or cross-reference is included.

TRN Trade Name of Product

Alphabetic characters.

TYPE Multum Medical Supply Category

Variable length alphabetic string.

Sample of Valid Values:

natural supplements
wound care supplies

UWT A semantic type provided from terms from the University of Washington Digital Anatomist

Repeating element.

Actual values:

Anatomical entity

Anatomical feature

Anatomical junction
Anatomical spatial entity
Anatomical structure
Body part
Body space
Body substance
Cell
Conceptual anatomical entity
Organ
Organ part
Organ system
Organ system subdivision
Physical anatomical entity
Tissue

VAC VA Class

The code of an NDF/HT drug class name.

Example:

COD LIVER OIL (CHERRY FLAVOR)
COD LIVER OIL, MINT FLAVORED

VID Health Level Seven Vocabulary ID

The numeric HL7 ID for an HL7/PT string.

Example:
1907

VMO VA CMOP (central mail-order pharmacy) ID

Alphanumeric string.

Example:

T0344
H0281

B.2. Vocabulary Source Abbreviations

AIR93
: AI/RHEUM. Bethesda (MD): National Library of Medicine, Lister Hill Center, 1993.

ALT2000
: Alternative Billing Concepts (AltLink). Version 983. Las Cruces (NM): Alternative Link LLC, 2000.

AOD99
: Alcohol and Other Drug Thesaurus: A Guide to Concepts and Terminology in Substance Abuse and Addiction. 3rd ed. [4 volumes]. Bethesda (MD): National Institute on Alcohol Abuse and Alcoholism (NIAAA) and Center for Substance Abuse Prevention (CSAP), 1999.

BI98
: Beth Israel OMR Clinical Problem List Vocabulary. Version 1.0. Boston (MA): Beth Israel Deaconess Medical Center, 1999.

BRMP2002
: Descritores em Ciencias da Saude [Portuguese translation of MeSH]. Sao Paulo (Brazil): Latin American and Caribbean Center on Health Sciences Information. BIREME/PAHO/WHO, 2002.

BRMS2002
: Descriptores en Ciencias de la Salud [Spanish translation of MeSH]. Sao Paulo (Brazil): Latin American and Caribbean Center on Health Sciences Information. BIREME/PAHO/WHO, 2002.

CCPSS99
: Canonical Clinical Problem Statement System (CCPSS). Version 1.0. Nashville (TN): Vanderbilt University, 1999.

CCS99
: Clinical Classifications Software (CCS). June 1999 release. Rockville (MD): Agency for Health Care Policy and Research (AHCPR), 1999.

CDT3
: Current Dental Terminology (CDT) contained in the HCFA Common Procedure Coding System (HCPCS). Version 3. Washington (DC): Health Care Financing Administration, 2002.

COS89
: Computer-Stored Ambulatory Records (COSTAR). Boston (MA): Massachusetts General Hospital, 1989. (List of terms that occur frequently at 3 COSTAR sites, supplied by Massachusetts General Hospital)

COS92
: Computer-Stored Ambulatory Records (COSTAR). Boston (MA): Massachusetts General Hospital, 1992. (List of terms that occur frequently at 3 COSTAR sites, supplied by Massachusetts General Hospital)

COS93
: Computer-Stored Ambulatory Records (COSTAR). Boston (MA): Massachusetts General Hospital, 1993. (List of terms that occur frequently at 3 COSTAR sites, supplied by Massachusetts General Hospital)

COS95
: Computer-Stored Ambulatory Records (COSTAR). Boston (MA): Massachusetts General Hospital, 1995. (List of terms that occur frequently at 3 COSTAR sites, supplied by Massachusetts General Hospital)

CPM93
: Columbia Presbyterian Medical Center Medical Entities Dictionary. New York (NY): Columbia Presbyterian Medical Center, 1993.

CPT2002
: Physicians' Current Procedural Terminology (CPT). 4th ed. Chicago (IL): American Medical Association, 2002.

CPT01SP
: Current Procedural Terminology (CPT), Spanish Translation. 4th ed. Chicago (IL): American Medical Association, 2001.

CSP2001
: Computer Retrieval of Information on Scientific Projects (CRISP). Bethesda (MD): National Institutes of Health, Division of Research Grants, Research Documentation Section, 2001.

CST95
: Coding Symbols for Thesaurus of Adverse Reaction Terms (COSTART). 5th ed. Rockville (MD): U.S. Food and Drug Administration, Center for Drug Evaluation and Research, 1995.

DDB00
: Diseases Database 2000. London (England): Medical Object Oriented Software Enterprises Ltd., 2000.

DMD2002
: German translation of MeSH. Cologne (Germany): Deutsches Institut fuer Medizinische Dokumentation und Information, 2002.

DMDICD
: Internationale Klassifikation der Krankheiten 10 [German translation of ICD10]. Cologne (Germany): Deutsches Institut fuer Medizinische Dokumentation und Information, 1998.

DMDUMD
: Die Nomenklatur fuer Medizinprodukte UMDNS [German translation of UMDNS]. Cologne (Germany): Deutsches Institut fuer Medizinische Dokumentation und Information, 2000.

DSM3R
: Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R). 3rd ed. rev. Washington (DC): American Psychiatric Association, 1987.

DSM4
Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Washington (DC): American Psychiatric Association, 1994.
DUT2001
Nederlandse vertaling van Mesh (Dutch translation of MeSH). Amsterdam, The Netherlands: Nederlands Tijdschrift voor Geneeskunde (Dutch Journal of Medicine), 2001.
DXP94
DXplain (An expert diagnosis program). Boston (MA): Massachusetts General Hospital.
FIN2002
Finnish translation of MeSH. Helsinki (Finland): Finnish Medical Society Duodecim, 2002.
HCDT3
HCPCS Version of Current Dental Terminology (CDT). Version 3. Washington (DC): Health Care Financing Administration, 2002.
HCPCS02
Healthcare Financing Administration Common Procedure Coding System (HCPCS). Washington (DC): Health Care Financing Administration, 2002.
HCPT02
HCPCS Version of Current Procedural Terminology (CPT). Washington (DC): Health Care Financing Administration, 2002.
HHC96
Saba, Virginia. Home Health Care Classification of Nursing Diagnoses and Interventions. Washington (DC): Georgetown University, 1996.
HL7
Health Level Seven Vocabulary (HL7). Ann Arbor (MI): Health Level Seven, 1998.
ICD10
International Statistical Classification of Diseases and Related Health Problems (ICD-10). 10th rev. Geneva (Switzerland): World Health Organization, 1998.
ICD10AE
International Statistical Classification of Diseases and Related Health Problems (ICD-10): Americanized Version. 10th rev. Geneva (Switzerland): World Health Organization, 1998.
ICD10AM
International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification; 2nd Edition, published January 2000.
ICD2002
The International Classification of Diseases: 9th revision, Clinical Modification (ICD-9-CM). 6th ed. Washington (DC): Health Care Financing Administration, July, 2001.
ICPC93
The International Classification of Primary Care (ICPC). Denmark: World Organisation of Family Doctors, 1993.

ICPC2E
: International Classification of Primary Care (ICPC) / prepared by the Classification Committee of the World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians (WONCA), known more briefly as the World Organization of Family Doctors. 2nd ed. Henk Lamberts and Inge Hofmans-Okkes, 1998.

ICPC2AE
: American English equivalent of the ICPC2E terms, produced by NLM. 2nd edition, 1998.

ICPC2P
: International Classification of Primary Care, Version 2-Plus, Australian Modification. January, 2000.

ICPCPAE
: American English equivalent of the ICPC2P terms, produced by NLM. Version 2-Plus, January, 2000.

ICPCBAQ
: Basque translation of ICPC93; see ICPC93

ICPCDAN
: Danish translation of ICPC93; see ICPC93

ICPCDUT
: Dutch translation of ICPC93; see ICPC93

ICPCFIN
: Finnish translation of ICPC93; see ICPC93

ICPCFRE
: French translation of ICPC93; see ICPC93

ICPCGER
: German translation of ICPC93; see ICPC93

ICPCHEB
: Hebrew translation of ICPC93; see ICPC93

ICPCHUN
: Hungarian translation of ICPC93; see ICPC93

ICPCITA
: Italian translation of ICPC93; see ICPC93

ICPCNOR
: Norwegian translation of ICPC93; see ICPC93

ICPCPOR
: Portuguese translation of ICPC93; see ICPC93

ICPCSPA
: Spanish translation of ICPC93; see ICPC93

ICPCSWE
: Swedish translation of ICPC93; see ICPC93

INS2001
: Thesaurus Biomedical Francais/Anglais [French translation of MeSH]. Paris (France): Institut National de la Sante et Recherche Medicale, 2001.

ITA2002
Italian translation of MeSH. Istituto Superiore di Sanita Servizio Documentazione, Viale Regina Elena, 229 00616 Rome, Italy; 2002.
JABL99
Online Congenital Multiple Anomaly/Mental Retardation Syndromes. Stanley Jablonski. Bethesda (MD): National Library of Medicine, 1999.
LCH90
Library of Congress Subject Headings. 12th ed. Washington (DC): Library of Congress, 1989.
LNC203
Logical Observation Identifiers, Names and Codes (LOINC). Version 2.03. Indianapolis (IN): The Regenstrief Institute, 2001.
MCM92
Glossary of Methodologic Terms for Clinical Epidemiologic Studies of Human Disorders. Canada: McMaster University.
MDDB99
First DataBank Master Drug Data Base (MDDB). San Bruno (CA): First DataBank Inc., 1999.
MDR40
Medical Dictionary for Regulatory Activities Terminology (MedDRA) Version 4.0. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), June 2001.
MDRAE40
Medical Dictionary for Regulatory Activities Terminology (MedDRA), American English Equivalents, Version 4.0. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), June 2001.
MDREA40
Medical Dictionary for Regulatory Activities Terminology (MedDRA), American English, with expanded abbreviations, Version 4.0. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), June 2001.
MDREX40
Medical Dictionary for Regulatory Activities Terminology (MedDRA),with expanded abbreviations, Version 4.0. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), June 2001.
MIM93
Online Mendelian Inheritance in Man (OMIM). Baltimore (MD): Johns Hopkins University, Center for Biotechnology Information, 1994.
MMSL01
Multum MediSource Lexicon. Denver (CO): Multum Information Services, Inc., July 2001.
MMX01
Micromedex DRUGDEX. Englewood (CO): Micromedex, 2001.

MSH2002
: Medical Subject Headings (MeSH). Bethesda (MD): National Library of Medicine, 2002.

MTH
: UMLS Metathesaurus. Bethesda (MD): National LIbrary of Medicine.

MTHCH02
: Metathesaurus Hierarchical CPT Terms (These terms were created by the NLM to provide contextual information for CPT). Bethesda (MD): National Library of Medicine.

MTHHH02
: Metathesaurus Hierarchical HCPCS Terms (These terms were created by the NLM to provide contextual information for HCPCS). Bethesda (MD): National Library of Medicine.

MTHICD9
: NLM-generated entry terms for ICD-9. Bethesda (MD): National Library of Medicine.

MTHMST2001
: Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy: International Edition April 22, 1998.

MTHMSTFRE
: Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy: French Edition April 22, 1998.

MTHMSTITA
: Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy: Italian Edition April 22, 1998.

NAN99
: Carroll-Johnson, Rose Mary, editor. Classification of Nursing Diagnoses. Proceedings of the 10th conference, 1999.

NCI2001a
: NCI Thesaurus. Version 3.0. Bethesda (MD): National Cancer Institute, National Institutes of Health, July 2001.

NDDF01
: National Drug Data File. San Bruno (CA): First DataBank Inc., 2001.

NEU99
: Bowden, Douglas M., Martin, Richard F., Dubach, Joev G. Neuronames Brain Hierarchy. Seattle (WA): University of Washington, Primate Information Center, 1999.

NIC99
: McCloskey, Joanne C., Bulechek, Gloria M., editors. Nursing interventions classification (NIC): Iowa intervention project. St. Louis (MO): Mosby-Year Book, 1999.

NOC97
: Johnson, Marion, Maas, Meridean, editors. Nursing Outcomes Classification (NOC): Iowa Outcomes Project. St. Louis (MO): Mosby-Year Book, 1997.

OMS94
Martin, Karen S., Scheet, Nancy J. The Omaha System: Applications for Community Health Nursing. Philadelphia (PA): W.B. Saunders, 1992 (with 1994 corrections).
PCDS97
Ozbolt, Judy Grace. Patient Care Data Set (PCDS). Version 4.0. Nashville (TN): Vanderbilt University School of Nursing, 1998.
PDQ2001
Physician Data Query Online System (PDQ). Bethesda (MD): National Cancer Institute, July 1 2001.
PPAC98
Pharmacy Practice Activity Classification (PPAC). Washington (DC): American Pharmaceutical Association. January, 1998.
PSY2001
Thesaurus of Psychological Index Terms. Ninth edition. Washington (DC): American Psychological Association, 2001.
QMR96
Quick Medical Reference (QMR). San Bruno (CA): First Databank, 1997.
RAM99
Randolph A. Miller Clinically Related Concepts. Nashville (TN): Vanderbilt University, 1999.
RCD99
Clinical Terms Version 3 (Read Codes) (Q199). England: National Health Service Centre for Coding and Classification, March, 1999.
RCDAE
American English equivalent of the Read Thesaurus terms produced by NLM. Version 3 (Q199). Bethesda (MD): National Library of Medicine, UMLS project, 1999.
RCDSA
American English equivalent of synthesized terms from the Read Thesaurus produced by NLM. Version 3 (Q199). Bethesda (MD): National Library of Medicine, UMLS project, 1999.
RCDSY
Synthesized Read terms (without initial bracketed letters) of the Read Thesaurus produced by NLM. Version 3 (Q199). Bethesda (MD): National Library of Medicine, UMLS project, 1999.
RUS2002
Russian Translation of MeSH. Moscow (Russia): State Central Scientific Medical Library, 2002.
SNM2
Cote, Roger A., editor. Systematized nomenclature of medicine. 2nd ed. Skokie (IL): College of American Pathologists, 1979. SNOMED update, 1982. Skokie (IL): College of American Pathologists, 1982.
SNMI98
Cote, Roger A., editor. Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International. Version 3.5. Northfield, (IL): College of

American Pathologists; Schaumburg (IL): American Veterinary Medical Association, 1998.

SPN99

Standard Product Nomenclature (SPN). Rockville (MD): U.S. Food and Drug Administration, 1999.

SRC

UMLS Metathesaurus Source Terminologies. Bethesda (MD): National Library of Medicine.

ULT93

Bell, Douglas. Ultrasound Structured Attribute Reporting (UltraSTAR). Boston (MA): Brigham & Womens Hospital, 1993.

UMD2002

Universal Medical Device Nomenclature System: Product Category Thesaurus. Plymouth Meeting (PA): ECRI, 2002.

UWDA155

University of Washington Digital Anatomist (UWDA). v. 1.55, Seattle (WA): University of Washington, January 2001.

VANDF01

U.S. Department of Veterans Affairs, Veterans Health Administration National Drug File. Department of Veterans Affairs, Washington, DC., September 5, 2001.

WHO97

WHO Adverse Drug Reaction Terminology (WHOART). Uppsala (Sweden): WHO Collaborating Centre for International Drug Monitoring, 1997.

WHOFRE

French translation of WHO97; see WHO97

WHOGER

German translation of WHO97; see WHO97

WHOPOR

Portuguese translation of WHO97; see WHO97

WHOSPA

Spanish translation of WHO97; see WHO97

B.2.1 Sources of additional (non-concept name) information

A small number of sources contribute information to the Metathesaurus but do not contribute concept names (i.e., the SAB does not appear in MRSO). For example, a source may contribute relationships between concepts, but not actually name the concepts. The following SABs do not appear in MRSO:

HDA99

Health devices alerts. Plymouth Meeting, PA: ECRI.

HLREL

ICPC2E-ICD10 relationships from Dr. Henk Lamberts (HLREL). University of Amsterdam. Contact: H.Lamberts@AMC.UVA.NL.

HPC99

Healthcare product comparison system. Plymouth Meeting, PA: ECRI.

MBD02

MEDLINE (1992-1996)

MED02

MEDLINE (1997-2001)

NCISEER

NCI Surveillance, Epidemiology, and End Results (SEER) conversions between ICD-9-CM and ICD-10 neoplasm codes. National Cancer Institute, Bethesda, MD. Release Date: June 1999. URL: http://www-seer.ims.nci.nih.gov/Admin/ConvProgs/ Phone: 301-496-8510.

NLM

National Library of Medicine (NLM). Bethesda (MD): National Library of Medicine.

OMIM97

Online Mendelian Inheritance in Man (OMIM). Bethesda (MD): National Center for Biotechnology Information, 1997. Contact: http://www3.ncbi.nlm.nih.gov/Omim.

B.3 Number of Strings From Each Source

AIR93	685
ALT2000	4878
AOD99	20669
BI98	1251
BRMP2002	42189
BRMS2002	40590
CCPSS99	15840
CCS99	1608
CDT3	508
COS89	776
COS92	735
COS93	626
COS95	1324
CPM93	536
CPT01SP	7894
CPT2002	16253
CSP2001	20261
CST95	6444

DDB00	256
DMD2002	48138
DMDICD	12003
DMDUMD	4449
DSM3R	467
DSM4	490
DUT2001	36495
DXP94	10113
FIN2002	20396
HCDT3	508
HCPCS02	4596
HCPT02	8336
HHC96	335
HL7	635
ICD10	13505
ICD10AE	1107
ICD10AM	25891
ICD2002	19860
ICDAMAE	2405
ICPC2AE	210
ICPC2E	3757
ICPC2P	13383
ICPC93	1053
ICPCBAQ	695
ICPCDAN	723
ICPCDUT	723
ICPCFIN	722
ICPCFRE	723
ICPCGER	723
ICPCHEB	485
ICPCHUN	718
ICPCITA	723

ICPCNOR	722
ICPCPOR	723
ICPCSPA	723
ICPCSWE	723
INS2001	29173
ITA2002	21194
JABL99	3260
LCH90	6652
LNC203	84604
MCM92	43
MDDB99	18239
MDR40	67427
MIM93	250
MMSL01	38789
MMX01	11539
MSH2002	488621
MTH	27792
MTHCH02	905
MTHHH02	323
MTHICD9	18357
MTHMST2001	1945
MTHMSTFRE	1833
MTHMSTITA	1799
NAN99	169
NCI2001a	2276
NDDF01	20088
NEU99	3796
NIC99	10187
NOC97	3056
OMS94	539
PCDS97	2229
PDQ2001	18778

PPAC98	380
PSY2001	7671
QMR96	943
RAM99	258
RCD99	347568
RCDAE	17315
RCDSA	1185
RCDSY	22186
RUS2002	42397
SNM2	44274
SNMI98	164180
SPN99	4851
SRC	321
ULT93	84
UMD2002	12373
UWDA155	79463
VANDF01	33972
WHO97	3831
WHOFRE	3717
WHOGER	3402
WHOPOR	3751
WHOSPA	3106

Total Counts

Concepts:	776,940
Terms:	1,601,727
Strings:	1,920,170
Source Strings:	2,104,784

B.4 Types of Names in a Vocabulary - the TTY

NOTE: In the following list, "term" is used to mean a name; all types of concept names are strings. This contrasts with the Metathesaurus definition of "term", which is the group of all strings that are lexical variants of each other.

AA	Attribute type abbreviation
AB	Abbreviation in any source vocabulary
AC	Activities
AD	Adjective
AS	Attribute type synonym
AT	Attribute type
BD	Fully-specified drug brand name that can be prescribed
BN	Fully-specified drug brand name that can not be prescribed
CC	Trimmed ICPC component process
CD	Clinical Drug
CE	Entry "term" to a Supplementary Concept "term"
CL	Class
CN	LOINC official component name
CO	ICPC component names (these are hierarchical terms, as opposed to the LOINC component names which are analytes)
CP	ICPC component process (in original form)
CS	Short component process in ICPC, i.e. include some abbreviations
CX	Component process in ICPC with abbreviations expanded
DE	Descriptor
DF	Dose Form
DFA	Dose Form Abbreviation
DI	Disease name
DO	Domain
DS	Short form of descriptor
DT	Definitional term, present in the Metathesaurus because of its connection to a Dorland's definition or to a definition created especially for the Metathesaurus
DX	Diagnosis
EN	MeSH nonprint entry "term"

EP	Entry "term"
ES	Short form of entry term
ET	Entry "term"
EX	Expanded form of entry term
FI	Finding name
FN	Full form of descriptor
GN	Generic drug name
GO	Goal
GQ	Geographic qualifier
GT	Glossary "term"
HC	Hierarchical class
HG	High Level Group Term
HS	Short hierarchical term (needed expansion) in ICD 10
HT	Hierarchical term
HX	Expanded version of short hierarchical term
ID	Nursing indicator
IN	Name for an intervention
IS	Obsolete synthesized term in the Read Thesaurus
IT	Index "term", i.e., derived from the index to any non-MeSH source vocabulary
IX	Expanded forms of indicators (embedded abbreviations expanded)
IV	Intervention
LN	LOINC official fully specified name
LO	Obsolete official fully specified name
LQ	Language qualifier.
LS	Expanded system/sample type (The expanded version was created for the Metathesaurus and includes the full name of some abbreviations.)
LT	Lower Level Term
LX	Official fully specified name with expanded abbreviations
MD	CCS multi-level diagnosis categories
MH	Main heading
MM	Metathesaurus string created to distinguish different meanings of the same lexical string

MP	Preferred names of modifiers
MS	Multum names of branded and generic supplies or supplements
MT	An alternate form of a concept name from one of the source vocabularies created for the Metathesaurus
MV	Multi-level procedure category
NM	Supplementary chemical "term", a name of a substance
NP	Non-preferred term
NS	Short form of non-preferred term
NX	Expanded form of non-preferred term
N1	Chemical Abstracts Service Type 1 name of a chemical
OA	Obsolete abbreviation
OC	Nursing outcomes
OL	Non-current Lower Level Term
OM	Obsolete modifiers in HCPCS
OP	Obsolete preferred term
OR	Orders
OS	System-organ class in the WHO Adverse Reaction Terminology
PC	Preferred "trimmed" term in ICPC
PM	Machine permutation
PN	Metathesaurus preferred name
PQ	Qualifier for a problem
PR	Name of a problem
PS	Short forms that needed full specification
PT	Designated preferred name
PX	Expanded preferred terms (pair with PS)
RN	Official component related name in LOINC
RS	Extracted related names in SNOMED2
RT	Designated related "term"
RX	Alternate name of preferred name
SA	Short forms of activities
SC	Special Category term
SCD	Semantic Clinical Drug
SCDC	Semantic Drug Component

SD	CCS single-level diagnosis categories
SF	Synonym made by replacing ";" with no spaces around it with ", " in ICPCP2
SI	Name of a sign or symptom of a problem
SN	Official component synonym in LOINC
SP	CCS single-level procedure categories
SS	Synonymous "short" forms
ST	Step
SX	Mixed-case component synonym with expanded abbreviations
SY	Designated synonym
TA	Task
TC	Term class
TG	Name of the target of an intervention
TQ	Topical qualifier
TT	ICD9CM table term
TX	CCPSS synthesized problems for TC termgroup
VS	Value Set
XD	Expanded descriptor in AOD
XQ	Alternate name for a qualifier

B.5 Order of Precedence of Source Concept Names as Distributed

NOTE: MTH/PN names always receive the highest ranking within a concept. After that, the ranking of an MTH/MM immediately precedes the highest ranking Source/Type of Name of the ambiguous string that caused it to be created. Also, please note that users are encouraged to change precedence to suit their own preferences.

The Metathesaurus file MRRANK contains this information which can be used to change precedence to suit your needs.

MTH/PN

MTH/MM

MSH2002/MH

MSH2002/HT

MSH2002/TQ

MSH2002/EP

MSH2002/EN

MSH2002/XQ

MSH2002/NM

DSM4/PT

DSM3R/PT

SNMI98/PT

SNMI98/PX

SNMI98/HT

SNMI98/HX

VANDF01/SCD

VANDF01/CD

VANDF01/HT

VANDF01/IN

VANDF01/SCDC

NDDF01/CD

NDDF01/IN

MDDB99/CD

MMX01/CD

MMX01/IN

RCDSA/PT

RCDSY/PT

RCDAE/PT

RCD99/PT

MSH2002/N1

MSH2002/CE

RCDSA/OP

RCDSY/OP

RCDAE/OP

RCD99/OP

SNM2/PT

SNMI98/RT

SNM2/RT

SNMI98/SY

SNMI98/SX

RCDSA/SY

RCDSY/SY

RCDAE/SY

RCD99/SY

RCDSA/IS

RCDSY/IS

RCDAE/IS

RCD99/IS

RCDAE/AT

RCD99/AT

RCD99/AS

SNMI98/AD

SNM2/SY

SNM2/RS

CPM93/PT

DDB00/PT

DDB00/SY

NEU99/HT

NEU99/PT

NEU99/SY

UWDA155/PT

UWDA155/SY

UMD2002/PT

UMD2002/ET

UMD2002/RT

MMSL01/CD

MMSL01/BD

MMSL01/SC

MMSL01/MS

MMSL01/GN

MMSL01/BN

MMSL01/IN

SPN99/PT

MDRAE40/HG

MDR40/HG

MDREA40/HG

MDREX40/HG

MDRAE40/PT

MDR40/PT

MDREA40/PT

MDREX40/PT

MDR40/OS

MDRAE40/HT

MDR40/HT

MDREA40/HT

MDRAE40/SC

MDREX40/HT

MDR40/SC

MDRAE40/LT

MDR40/LT

MDREA40/LT

MDREX40/LT

CST95/PT

WHO97/OS

WHO97/HT

WHO97/PT

WHO97/IT

AIR93/HT

AIR93/FI

AIR93/DI

AIR93/SY

ULT93/PT

CPT2002/PT

CPT2002/SY

CPT2002/MP

HCPT02/PT

HCPCS02/PT

CDT3/PT

HCDT3/PT

HCPCS02/MP

HCPT02/MP

ICD10AE/PT

ICD10/PT

ICD10AE/PX

ICD10/PX

ICD10AE/PS

ICD10/PS

ICDAMAE/PT

ICD10AM/PT

ICDAMAE/PX

ICD10AM/PX

ICDAMAE/PS

ICD10AM/PS

PDQ2001/PT

PDQ2001/SY

NCI2001a/PT

NCI2001a/SY

NCI2001a/AB

ICPC2AE/PT

ICPC2E/PT

ICPC2AE/PX

ICPC2E/PX

ICPC93/PX

ICPC93/PT

ICPC2AE/PS

ICPC2E/PS

ICPC2AE/ET

ICPC2E/ET

ICPC93/PS

ICPC93/PC

ICPC93/CX

ICPC93/CP

ICPC93/CS

ICPC93/CC

ICPC2E/CO

ICPC93/CO

ICPC2AE/AB

ICPC2E/AB

CCPSS99/TX

CCPSS99/TC

CCPSS99/PT

CCPSS99/MP

ICPCPAE/SF

ICPCPAE/SY

ICPC2P/SF

ICPC2P/SY

ICPCPAE/PX

ICPC2P/PX

ICPCPAE/PT

ICPC2P/PT

ICPCPAE/PS

ICPC2P/PS

AOD99/DE

AOD99/DS

AOD99/XD

AOD99/FN

AOD99/ET

AOD99/ES

AOD99/EX

AOD99/NP

AOD99/NS

AOD99/NX

HCPCS02/OP

CDT3/OP

HCDT3/OP

HCPT02/OP

HCPCS02/OM

HCPT02/OM

JABL99/PC

JABL99/PT

JABL99/SS

JABL99/SY

MIM93/PT

PDQ2001/RT

BI98/PT

BI98/SY

BI98/RT

LNC203/LX

LNC203/LN

LNC203/LO

LNC203/CX

LNC203/CN

LNC203/SX

LNC203/SN

LNC203/LS

LNC203/RX

LNC203/RN

DSM4/HT

DSM3R/HT

SNM2/HT

ICD2002/PT

MDRAE40/OL

MDR40/OL

MDREX40/OL

ICD2002/HT

CCS99/HT

CCS99/MD

CCS99/SD

CCS99/MV

CCS99/SP

ICD10AE/HT

ICD10/HT

ICD10AE/HX

ICD10/HX

ICD10AE/HS

ICD10/HS

ICDAMAE/HT

ICD10AM/HT

UMD2002/HT

ICPC93/HT

RAM99/PT

RAM99/RT

QMR96/PT

HL7/PT

HL7/DF

HL7/DFA

HL7/VS

MTHCH02/HT

MTHHH02/HT

HHC96/DX

BI98/AB

HHC96/IV

HHC96/CO

NIC99/IV

NIC99/HC

NAN99/PT

NAN99/HT

NAN99/RT

OMS94/MT

OMS94/PR

OMS94/TG

OMS94/HT

OMS94/PQ

OMS94/IV

OMS94/SI

NIC99/AC

NIC99/SA

NOC97/OC

NOC97/IX

NOC97/ID

PCDS97/GO

PCDS97/OR

PCDS97/PR

NIC99/HT

NOC97/HT

NOC97/HC

HHC96/MP

PCDS97/CO

PCDS97/HX

PCDS97/HT

COS95/PT

COS93/PT

COS92/PT

COS89/PT

DXP94/DI

DXP94/FI

DXP94/SY

MCM92/PT

MCM92/RT

PPAC98/DO

PPAC98/CL

PPAC98/AC

PPAC98/ST

PPAC98/TA

ALT2000/PT

ALT2000/SY

ALT2000/HT

MTH/PT

MTH/SY

MTH/RT

DSM3R/SY

DSM3R/RT

MTHICD9/ET

CST95/SC

CST95/HT

CST95/GT

PSY2001/PT

PSY2001/HT

PSY2001/ET

MTHMST2001/PT

MTHMST2001/SY

LCH90/PT

MSH2002/PM

RCDSA/AB

RCDSY/AB

RCDAE/AB

RCD99/AB

RCDSA/OA

RCDSY/OA

RCDAE/OA

RCD99/OA

RCDAE/AA

RCD99/AA

CSP2001/PT

CSP2001/SY

CSP2001/ET

MTH/DT

BRMP2002/MH

BRMS2002/MH

DUT2001/MH

DMD2002/MH

FIN2002/MH

FIN2002/EP

INS2001/MH

ITA2002/MH

RUS2002/MH

BRMP2002/SY

BRMS2002/SY

DUT2001/SY

DMD2002/SY

INS2001/SY

ITA2002/SY

RUS2002/SY

BRMP2002/EP

BRMS2002/EP

DUT2001/EP

DMD2002/EP

DMDUMD/PT

DMDUMD/ET

DMDUMD/RT

WHOFRE/OS

WHOGER/OS

WHOPOR/OS

WHOSPA/OS

WHOFRE/HT

WHOGER/HT

WHOPOR/HT

WHOSPA/HT

WHOFRE/PT

WHOGER/PT

WHOPOR/PT

WHOSPA/PT

WHOFRE/IT

WHOGER/IT

WHOPOR/IT

WHOSPA/IT

CPT01SP/PT

DMDICD/PT

DMDICD/HT

ICPCBAQ/PT

ICPCDAN/PT

ICPCDUT/PT

ICPCFIN/PT

ICPCFRE/PT

ICPCGER/PT

ICPCHEB/PT

ICPCHUN/PT

ICPCITA/PT

ICPCNOR/PT

ICPCPOR/PT

ICPCSPA/PT

ICPCSWE/PT

ICPCBAQ/CP

ICPCDAN/CP

ICPCDUT/CP

ICPCFIN/CP

ICPCFRE/CP

ICPCGER/CP

ICPCHEB/CP

ICPCHUN/CP

ICPCITA/CP

ICPCNOR/CP

ICPCPOR/CP

ICPCSPA/CP

ICPCSWE/CP

MTHMSTFRE/PT

MTHMSTITA/PT

SRC/PT

SRC/SY

SRC/HT

SRC/AB

B.6 Source Specific Relationship Attributes Not Listed in the Semantic Network

adjectival_form_of

classified_as

classifies

clinically_associated_with

clinically_similar

closest_related_pdq_diagnosis

ddx - A relationship attribute that represents Differential Diagnosis

default_mapped_from

default_mapped_to

dose_form_of

equivalent_to

has_closest_related_pdq_diagnosis

has_dose_form

has_member

icd_asterisk

icd_dagger

mapped_from

mapped_to

member_of_cluster

multiply_mapped_from

multiply_mapped_to

noun_form_of

other_mapped_from

other_mapped_to

primary_mapped_from

primary_mapped_to

sib_in_branch_of

sib_in_isa

sib_in_part_of

sib_in_tributary_of

similar

ssc - A relationship attribute that represents Sign Symptom Complex

uniquely_mapped_from

uniquely_mapped_to