
Filtering the UMLS[®] Metathesaurus[®] for MetaMap

2007 Edition

Willie J. Rogers and Alan R. Aronson

January 4, 2008

1. Overview

The MetaMap program's purpose is to discover the Metathesaurus concepts referred to in arbitrary text. A given Metathesaurus concept can have many alternative names (Metathesaurus strings) which originate in the many source vocabularies included in the Metathesaurus. As the number of strings has grown over the years, MetaMap's performance has suffered. In the 2007AA version of the Metathesaurus, for example, the Metathesaurus has 3,331,382 English strings, 3,301,112 (99.1%) of them distinct, comprising 1,359,705 concepts. There are 8% more English strings and 7.6% more concepts than in the 2006AA edition. Many of the strings in the Metathesaurus are of little value to MetaMap for one of four reasons. First, some strings either represent general, nonmedical concepts or are unnecessarily ambiguous. Second, some strings are virtually indistinguishable from each other. For efficiency, only one representative of a set of indistinguishable strings is needed. Third, some Metathesaurus strings have internal structure or meta information which makes it highly unlikely to find them in regular text. Fourth, some strings, including lengthy descriptions of things such as procedures, health activities or medical devices, are so complicated that it is again unlikely to find them in normal text. Corresponding to the four classes of strings are four filtering methods for discovering and removing them: manual filtering, lexical filtering, filtering by type, and syntactic filtering. These methods are discussed in sections 2-5. Then section 6 describes ways to selectively combine the filtering methods to produce a range of alternative views of the Metathesaurus appropriate for various purposes. Finally, section 7, is an Appendix providing details of filtering decisions for Metathesaurus strings based on term type (TTY).

2. Manual Filtering

A number of Metathesaurus strings are problematic and have been suppressed before performing other forms of filtering. There are 11,621 such strings, 9% more than in 2006 (There were significant increases in Unnecessarily ambiguous terms, Brand Names, and Numeric terms):

- **Unnecessarily ambiguous terms** [3,337 occurrences]
 - ‘Other’ for ‘Other location of complaint’
 - ‘Protocols’ for ‘Protocols: Urinary Elimination’
 - ...
- **Contextual terms**, i.e., terms whose meaning can only be understood within the context of their vocabulary [7,131 occurrences]
 - All terms containing “*NEC*” or an expanded form
- **Brand names**, i.e., short forms of terms containing “*brand*” [340 occurrences]
- **Enzyme Commission (EC) numbers** beginning “*EC <integer>*.” [161 occurrences]
- **Single alphabetic strings** (e.g., ‘a’, ‘A’, ‘b’, ‘B’) [157 occurrences];
- **Numbers** (e.g., ‘2’, ‘+1’, ‘-4’, ‘98.734’, ‘50000’) [450 occurrences];
- **Special cases** [5 occurrences]
 - ‘Periods’ and ‘Period’ for ‘Menstruation’ (C0025344)
 - ‘Clap’ and ‘CLAP’ for ‘Gonorrhoea’ (C0018081)
 - ‘BRA’ for ‘Brain’ (C0006104)

The Metathesaurus staff marks some terms as suppressible synonyms because they are thought to be inappropriate for any use. These are terms that, for one reason or another, do not adequately describe the concept that contains them. They are given a Term Status (TS) of lowercase s (or p) and are discussed in the annual editions of *Ambiguity in the UMLS Metathesaurus*. Contextual terms are actually a specific kind of ambiguous term. They are identifiable by the presence of *NEC* or one of its expansions (e.g., not elsewhere classified). Brand names are problematic because they often consist of a common word (e.g., *Cold*) which almost never has the brand name meaning in biomedical text. Although a few numbers correspond to biomedical entities (“98.734” has semantic types ‘Steroid’ and ‘Pharmacologic Substance’), they generally have semantic types ‘Quantitative Concept’ or ‘Intellectual Product’. Similarly, the single alphabetic generally mean the letter itself (the concept for “a” is “Lower case ay”) and have semantic type ‘Intellectual Product’. Several single alphabetic, however are biomedical: “B” has concept “Boron” with semantic type ‘Element, Ion, or Isotope’. A final class of special cases includes the string “Periods” for “Menstruation” and “BRA” for “Brain”. Both of these are problematic because they are ambiguous with other concepts which occur far more frequently in biomedical text.

3. Lexical Filtering

Lexical filtering is the most benign type of filtering and consists of removing strings for a concept which are effectively the same as another string for the concept. Properties which can make strings effectively the same are:

- non-essential parentheticals;
- *NOS* variation;

- syntactic uninversion;
- case variation;
- hyphen variation; and
- possessives.

Lexical filtering is accomplished by normalizing all strings for a given concept and removing all but one string for each set of strings that normalize to the same thing.

3.1 Non-essential parentheticals

Non-essential parentheticals are parenthetical expressions within a string which provide meta information about the string. As such they are not useful for text processing. Non-essential parentheticals can occur at the left or right end of a string and can be delimited by either parentheses or brackets. For example the concept “Anemia, Hemolytic” has synonyms “[X]Haemolytic anaemias” and “[X]Hemolytic anemias” both of which contain the left parenthetical *[X]*. Previous editions of the Metathesaurus only contained right parentheticals which seemed to be relatively well-behaved in the sense that a string without the parenthetical was almost always present in the set of strings for a given concept. Thus, “Drug Toxicity (Non MeSH)” had a string “Drug Toxicity”. Now right parentheticals are much less well-behaved and only a few left parentheticals can be reliably removed without altering the string’s meaning. These left parentheticals come from the Read Codes (and also SNOMEDCT): *[X]*, *[V]*, *[D]*, *[M]*, *[EDTA]*, *[SO]* and *[Q]*. These are the only parentheticals declared to be non-essential and removed from strings. The problem of detecting non-essential parentheticals has changed as the Metathesaurus has matured. The current practice of removing the few left parentheticals listed above is by no means adequate. The problem requires further analysis.

3.2 NOS variation

Many of the Metathesaurus vocabularies incorporate the acronym *NOS* (*Not Otherwise Specified*) into their terms. Examples include “Abdomen, NOS” and “X-RAY NEC AND NOS”. As with case variation, the presence of *NOS* (except when also accompanied by *NEC*) does not generally have a significant effect on the meaning of the term. The argument for ignoring *NOS* variation is not as strong as that for case variation, but it still seems reasonable for most text processing.

3.3 Syntactic uninversion

Inversion refers to the practice of inverting words of a term and inserting a comma to signal the inversion. It is normally done to index the original term under each of its important words and thereby make it more accessible. Inverted forms of a term, however, are not useful for processing text since inverted forms rarely appear in text. The concept “1,4-alpha-Glucan Branching Enzyme” has some interesting inversions. It has a synonym “Branching Enzyme” with inversion “Enzyme, Branching”, and it also has a synonym “Starch Branching Enzyme” with two inversions, “Branching Enzyme, Starch” and “Enzyme, Starch Branching”. The process of uninversion simply undoes inversion, i.e., it searches for a comma followed by a space, inverts the term at that point and removes the comma and space. Syntactic uninversion is just uninversion which is inhibited if the term contains a preposition or conjunction. This prevents terms such as “Biological Phenomena, Cell Phenomena, and Immunity” or “Legal blindness, as defined in U.S.A.” from being incorrectly uninverted. Note that the concept “1,4-alpha-Glucan Branching Enzyme” men-

tioned earlier is also not uninverted because the comma within it is not followed by a space; such embedded commas do not call for uninversion.

3.4 Case variation

Two strings which differ from each other only because of case variation normally refer to the same thing. For example, the concept ‘Abdomen’ has strings “abdomen” and “ABDOMEN” in addition to “Abdomen” which differ from each other only by case. Similarly, the concept ‘beta-Alanine’ has strings, “beta Alanine”, “beta alanine” and “BETA ALANINE”, which differ from each other only by case. Note, however, that case *does* matter for some aspects of text processing. Text containing the pronoun *us* is not referring to the acronym *US* for the *United States*; and the verb *aids* does not refer to the disease *AIDS*. Similarly, the case variation in the first three letters of the concepts ‘CDE genotype’, ‘CDe genotype’, ..., and ‘cde genotype’ is significant. Despite these observations, case almost never matters within the limited context of the set of all strings for a given concept.

3.5 Hyphen variation

As with case variation, the presence of a hyphen rather than a space normally means little especially in the context of all strings for a given concept. For example, the concept “1,4-alpha-Glucan Branching Enzyme” used in the last section has a variant “1,4 alpha Glucan Branching Enzyme” in which both hyphens have been replaced by spaces.

3.6 Possessives

Alternatives such as “Down’s Syndrome” and “Down Syndrome” or “American Nurses’ Association” and “American Nurses Association” differ only by a possessive.

4. Filtering by Type

Some Metathesaurus strings can be filtered out based solely on their type. For example, strings with a Term Status (TS) of lowercase s or p are suppressible synonyms or suppressible preferred names (see section 2 above). “Abdomen” is such a synonym of “Malignant neoplasm of abdomen.” Almost 11% (370,393) of the 3,331,382 English Metathesaurus strings are marked as suppressible synonyms. There is also a smaller number (69,921) of strings with TS of lowercase p.

Similarly, some Term Type (TTY) values indicate strings which are normally inappropriate for text processing, often because they are abbreviatory in nature, have some internal structure or include some meta information (see section 7.1 below). There are 1,118,143 occurrences of strings with such term types in the Metathesaurus. Actually there is a non-trivial overlap between these inappropriate term types and the suppressible synonyms: 349,616 (31%, down from 35% in 2006) of the 1,118,143 TTY strings are already marked as suppressible.

The filtered out term types together with examples for each are given in the Appendix (section 7). For completeness, the Appendix also includes a section of questionable types and a section of good types.

5. Syntactic Filtering

The final kind of filtering considered here is based on a high-level syntactic parse of the Metathesaurus strings. Since normal MetaMap processing involves mapping the simple noun phrases found in text, it is highly unlikely that a complex Metathesaurus string will be part of a good mapping. For example, the concept “Accident caused by caustic and corrosive substances” has high-level syntactic analysis [[head],[verb],[prep,head],[conj],[mod,head]] which contains seven syntactic units (head, verb, etc.) broken into five simple phrases ([head], [verb], etc.) Any text which resembles the concept will be broken up into several phrases each of which is processed separately. Thus, the text might map to constituent concepts (such as “Accident”); but the entire text will not map to the full concept. The strictest form of syntactic filtering, then, would be to filter out any string consisting of more than one simple phrase. However some tractable strings with more than one simple phrase are not filtered out. As of 1999, for example, *of strings* such as “Acute necrosis of liver” and “Radical resection of tumor of soft tissue of leg area”, which consist of a simple phrase followed by one or more *of* prepositional phrases, have not been excluded in syntactic filtering because of their tractability. In 2001 this condition was relaxed further to include phrases consisting of a simple phrase followed by any prepositional phrase followed by zero or more *of* prepositional phrases. An example of such a phrase is “Other operations on vessels of heart”.

6. Filtered Metathesaurus Models

The filtering described in the previous sections can be selectively applied to provide different views of the Metathesaurus. Three such models are

- **Strict Model:** All forms of filtering, manual, lexical, type-based and syntactic, are applied. This view is most appropriate for semantic processing where the highest level of accuracy is needed. The Strict Model consists of 1,937,745 (58%) of the 3,331,382 English Metathesaurus strings;
- **Moderate Model:** Manual, lexical and type-based filtering, but not syntactic filtering, are used. This view is appropriate for term processing where input text should not be divided into simple phrases but considered as a whole. The Moderate Model consists of 2,199,211 (66%) English Metathesaurus strings; and
- **Relaxed Model:** Only manual and lexical filtering are performed. This provides access to virtually all Metathesaurus strings and is appropriate for browsing. The Relaxed Model consists of 3,051,888 (92%) English Metathesaurus strings.

Note that in order to have all Metathesaurus concepts represented in each model, the preferred name of a concept is retained when all of its strings would otherwise be filtered out. For the strict model 458,281 concepts are so spared, for the moderate model 23,159 are spared, and for the relaxed model no concepts need to be spared.

7. Appendix

7.1 Filtered out Types

- AA (Attribute type abbreviation) [34 occurrences]
 - “Route administration of drug” (for concept “Drug Administration Routes”)
 - “Type-partial denture connector” (for concept “Type of partial denture connector”)
- AB (Abbreviation in any source vocabulary) [85,318 occurrences]
 - “Cluster of diff antigen 73” (for term “Cluster of differentiation antigen 73” of concept “5’-Nucleotidase”)
 - “AIDS dementia complex” (for concept “AIDS Dementia Complex”; note that in this case the AB string is not unusual, but it is redundant)
- ACR (Acronym) [3,425 occurrences]
 - “FA”
 - “BLEO/DOX/DTIC/VCR”
- CCN (Chemical code name) [1,096 occurrences]
 - “EC 3.5.1.1”
 - “T-1220”
- CCS (FIPS 10-4 country code) [237 occurrences]
 - “AF”
 - “AL”
- CDA (Clinical drug name in abbreviated format) [20,873 occurrences]
 - “BEESWAX MISCELL. WAX (GM)”
 - “CASTOR OIL MISCELL. OIL”
- CDC (Clinical drug name in concatenated format (NDDF), Clinical drug name (NDFRT)) [30,412 occurrences]
 - “ALCOHOL %”
 - “BEESWAX MISCELL. WAX (GM)”
- CDD (Clinical drug name in delimited format) [20,873 occurrences]
 - “BEESWAX@@MISCELL.@WAX (GM)”
 - “CASTOR OIL@@MISCELL.@OIL”
- CO (ICPC component names (these are hierarchical terms, as opposed to the LOINC component names which are analytes)) [57 occurrences]
 - Like topical qualifiers (TQ), these terms seem to have a meaning too specific to be useful for most biomedical applications
 - “ACTIVITY COMPONENT”
 - “CARDIAC COMPONENT”

-
- CS (Short component process in ICPC, i.e. include some abbreviations) [13 occurrences]
“Med exam/health evalua/complete”
“Microbio/other immunol test”
 - CSN (HL7 Codesystem Name) [2,396 occurrences]
“HL7V3.0”
“Confidentiality”
 - FN (Full form of descriptor) [307,076 occurrences]
“Methylphenyltetrahydropyridine (substance)”
“Acute abdomen (disorder)”
 - HS (Short hierarchical term (needed expansion) in ICD 10) [197 occurrences]
“Agents primarily acting on smooth and skeletal muscles and the respiratory system”
“Bacterial vaccines”
 - HTN (HL7 Table Name) [478 occurrences]
“Bed”
“Artificial blood”
 - HX (Expanded version of short hierarchical term) [5,577 occurrences]
“D40-D44 ABDOMEN” (for concept “Abdomen”; the HX form occurs in SNMI98 and has an HT form “ABDOMEN”)
“SECTION 2 CONGENITAL ANOMALIES” (for concept “Congenital Abnormality”; the HX form occurs in SNMI98 and has an HT form “CONGENITAL ANOMALIES”)
 - JAXPT (NCI Mouse Terminology PT) [138 occurrences]
“AKR”
“BALB/c”
 - JAXSY (NCI Mouse Terminology SY) [18 occurrences]
“severe combined immune deficiency”
“C3Smn.C-Prkdc-scid/J”
 - LN (LOINC official fully specified name) [44,322 occurrences]
“ALMECILLIN:SUSC:PT:ISLT:ORDQN:MIC”
“ALMECILLIN:SUSC:PT:ISLT:ORDQN:AGAR DIFFUSION”
 - LO (Obsolete official fully specified name) [1,216 occurrences]
“AUREOBASIDIUM PULLULANS AB.IGE:ACNC:PT:SER:QN”
“PARROT AUSTRALIAN DROPPINGS AB.IGE:ACNC:PT:SER:QN”
 - LX (Official fully specified name with expanded abbreviations) [44,268 occurrences]
“ALMECILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE:QUANTITATIVE:MINIMUM INHIBITORY CONCENTRATION”

“ALMECILLIN:SUSCEPTIBILITY:POINT IN TIME:ISOLATE:SEMI-QUANTITATIVE:BACTERIAL SENSITIVITY (KIRBY-BAUER)”

- MTH_IS (Metathesaurus-supplied form of obsolete synthesized term in the Read Thesaurus) [114 occurrences]
 - “24,25-Dihydroxyvitamin D₃”
 - “beta⁺ Thalassemia, NOS”
- MTH_OES (Metathesaurus obsolete exact synonym expanded) [3 occurrences]
 - “dihydrolipoamide dehydrogenase complex location”
 - “holo-ACP synthase complex location”
- MTH_OF (Metathesaurus-supplied form of obsolete fully specified name) [261 occurrences]
 - “Vitamin B₄ (substance)”
 - “Vitamin B4 (substance)”
- MTH_OP (Metathesaurus obsolete preferred term) [542 occurrences]
 - “Vitamin B4”
 - “Vitamin B₄”
- MTH_OPN (Metathesaurus obsolete preferred term, natural language form) [8 occurrences]
 - “megaloblastic anemia”
 - “Barretts esophagitis”
- MTH_ORIS (Metathesaurus obsolete related synonym expanded) [1 occurrence]
 - “proteasome endopeptidase complex location”
- MTH_SC (MTH Special Category term) [3 occurrences]
 - “CARDIAC ISCHEMIA (SSC)”
 - “EDEMA (SSC)”
- NPT (HL7 non-preferred for language term) [171 occurrences]
 - “Audiologist”
 - “flavor”
 - “Aleut”
- OA (Obsolete abbreviation) [49,131 occurrences]
 - “Gastrin secretion abnorm.NOS”
 - “Spontaneous abort.incomp.NOS”
- OB (Obsolete term) [1,136 occurrences]
 - “Acetaminophen”
 - “Paracetamol”
- OCD (Obsolete clinical drug) [224,387 occurrences]
 - “BEESWAX MISC. WAX (GM)”

“CASTOR OIL MISC. OIL”

“Adhesive Tape”

“Glucose Oral Soln 50%”

Note that as of 2008, OCD and OF will be classified as questionable types because the great majority of their strings which are not appropriate for text processing are already marked as suppressible.

- OES (Obsolete exact synonym) [254 occurrences]
 - “muscle fibre”
 - “CstF”
- OF (Obsolete fully specified name) [113,756 occurrences]
 - “Dipalmitoylphosphatidylcholine (substance)”
 - “17-Hydroxycorticosteroid -RETIRED-”

Note that as of 2008, OCD and OF will be classified as questionable types because the great majority of their strings which are not appropriate for text processing are already marked as suppressible.
- OLX (Expanded LOINC obsolete fully specified name) [1,215 occurrences]
 - “CRYPTOSPORIDIUM:::”
 - “5-FLUOROCYTOSINE:SUSCEPTIBILITY:POINT IN TIME:ISOLATE:QUANTITATIVE OR ORDINAL:MINIMUM INHIBITORY CONCENTRATION”
- OOSN (Obsolete official short name) [1,167 occurrences]
 - “Deprecated ^”
 - “Deprecated 5FC MIC”
- PS (Short forms that needed full specification) [1,740 occurrences]
 - “Cranial nerves” (for concept “Benign neoplasm of cranial nerves”)
 - “Bladder, unspecified” (for concept “Malignant neoplasm of bladder, NOS”)
- RAB (Root abbreviation) [131 occurrences]
 - “DSM3R”
 - “MSH”
- RHT (Root hierarchical term) [54 occurrences]
 - “DSM-III-R”
 - “MeSH”
- RPT (Root preferred term) [131 occurrences]
 - “DSM-III-R”
 - “Medical Subject Headings”
- RSY (Root synonym) [35 occurrences]
 - “Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R”
 - “LCSH”

-
- SB (Named subset of a source) [1 occurrence]
“US English Dialect Subset”
 - SSN (Source short name, used in the UMLS Knowledge Source Server) [131 occurrences]
“DSM-III-R”
“MeSH”
 - UCN (Unique common name) [203 occurrences]
“canaries <genus>”
“monkeys <#2>”
 - USN (Unique scientific name) [1,072 occurrences]
“Absidia <zygomycete fungus>”
“Aedes <sugbenus>”
“Bacillus <bacterium>”
 - USY (Unique synonym) [90 occurrences]
“Chlamydia psittaci <Chlamydophila psittaci>”
“AsGV<agrotis>”
 - VAB (Versioned abbreviation) [128 occurrences]
“SPN2003”
“GO2004_12_20”
“MSH2006_2005_11_15”
 - VPT (Versioned preferred term) [128 occurrences]
“Standard Product Nomenclature, 2003”
“Medical Subject Headings, 2006_2005_11_15”
 - VSY (Versioned synonym) [17 occurrences]
“LCSH, 1990”
“Glossary of Methodologic Terms for Clinical Epidemiologic Studies of Human Disorders, 1992”
 - XM (Cross mapping set) [9 occurrences]
“MSH Associated Expressions”
“SNOMEDCT mappings to ICD-9-CM”
 - XX (Expanded string) [69 occurrences]
“superior frontal sulcus (human only)”
“ectocalcarine sulcus (macaque only)”

7.2 Questionable Types

The Term Types listed in this section are not actually filtered out during type filtering, but they present problems for processing text adequately.

-
- BD (Fully-specified drug brand name that can be prescribed) [19,001 occurrences]
These terms describe a quantity of some drug; they may or may not occur in text.
“Adalat, 10 mg oral capsule”
“Afrin, 0.05% nasal spray”
 - CC (Trimmed ICPC component process) [1 occurrence]
“Referral primary care provider”
 - CD (Clinical Drug) [118,016 occurrences]
Like BD terms, these terms often describe a quantity of some drug and may or may not occur in text.
“Lavender Oil”
“Hydrogen Peroxide Soln 3%”
“Isopropyl Alcohol 99%”
 - CDISCPY (Clinical Data Interchange Standards Consortium PT) [731 occurrences]
“TZA”
“TRANSDERMAL”
“SURGICAL AND MEDICAL PROCEDURES”
 - CDISCSY (Clinical Data Interchange Standards Consortium SY) [387 occurrences]
“BELIZE”
“MYANMAR”
“Medical Dictionary for Regulatory Activities”
 - CP (ICPC component process (in original form)) [25 occurrences]
“Administrative procedure”
“Other therapeutic procedure”
 - CV (Content View) [1 occurrence]
“MetaMap NLP View”
 - DS (Short form of descriptor) [509 occurrences]
“AOD withdrawal syndrome”
“biological AOD dependence”
 - EQ (Equivalent name) [2,784 occurrences]
“Borrelia burgdorferi sensu stricto”
“[Prevotella] zoogloformans”
 - ES (Short form of entry “term”) [36 occurrences]
“periodic light AOD use”
“AOD tax”
 - GO (Goal) [311 occurrences]
Like orders (OR) below, the terms are full utterances.

“Mobility, exercise, and activity will increase to optimal or return to baseline.”

“Patient’s activity tolerance will increase or progress.”

- HD (Hierarchical descriptor) [122 occurrences]
 - “Transplanted Organ Complication”
 - “Neoplasm by Site”
- IS (Obsolete synthesized “term” in the Read Thesaurus) [70,350 occurrences]
 - “Fall - accidental” (for concept “Accidental Falls”)
 - “MVTA - Motor vehicle traffic accident” (for concept “Accidents, Traffic”)
- IX (Expanded forms of indicators (embedded abbreviations expanded)) [257 occurrences]
 - “Systolic blood pressure”
 - “Exercise stress test within normal limits”
- LPDN (LOINC parts display name) [101 occurrences]
 - “Occupational Allergen Mix 2 (Acarus siro + Lepidoglyphus destructor + Gasterophilus intestinalis + Sitophilus granarius)”
 - “Tree Allergen Mix 8 (Boxelder + Silver birch + Hazelnut + White oak + London plane/Maple leaf sycamore)”
- LS (Expanded system/sample type (The expanded version was created for the Metathesaurus and includes the full name of some abbreviations.)) [1,646 occurrences]

About one-third of these terms contain embedded periods.

 - “AORTIC VALVE”
 - “AORTA.THORACIC.ASCENDING”
- MP (Preferred names of modifiers) [197 occurrences]
 - “ABNORMALITY”
 - “PROB”
- MT (An alternate form of a concept name from one of the source vocabularies created for the Metathesaurus) [104 occurrences]
 - “coma”
 - “incontinence of stool”
- MTH_FN (MTH Full form of descriptor) [3,088 occurrences]
 - “2,4,5-trichloroacetophenol (substance)”
 - “Adjustment disorder with disturbance of conduct (disorder)”
- MTH_OL (MTH Non-current Lower Level Term) [131 occurrences]
 - “Spinal anesthesia (all forms)”
 - “Respiratory arrest (excluding neonatal)”

-
- MTH_PTGB (Metathesaurus-supplied form of British preferred term) [138 occurrences]
“delta beta⁰ Thalassaemia”
“Haemoglobin A₂ Indonesia”
 - MTH_SY (MTH Designated synonym) [830 occurrences]
“25-Hydroxyvitamin D₂”
“Bacterial Ca²⁺/Mg²⁺ ATPase”
 - MTH_SYGB (Metathesaurus-supplied form of British synonym) [19 occurrences]
“Hereditary persistence of fetal haemoglobin (HPFH) G gamma beta⁺ thalassaemia”
“Vitamin B₁₂ deficiency anaemia”
 - MV (Multi-level procedure category) [403 occurrences]
“Adenoidectomy without tonsillectomy”
“Excision of semilunar cartilage of knee”
 - NS (Short form of non-preferred “term”) [200 occurrences]
“neonatal AOD abstinence syndrome”
“dysfunctional AOD use”
 - NX (Expanded form of non-preferred “term”) [200 occurrences]
“neonatal Alcohol or Other Drugs abstinence syndrome”
“dysfunctional Alcohol or Other Drugs use”
 - OC (Nursing outcomes) [193 occurrences]
“Thermoregulation”
“Decision Making”
 - OL (Non-current Lower Level Term) [8,999 occurrences]
“Anomaly anomaly congen”
“Congenital anomaly, unspecified”
 - ONP (Obsolete non-preferred for language term) [37 occurrences]
“Retired Code”
 - ONS (Obsolete narrow synonym) [10 occurrences]
“germination (sensu Saccharomyces)”
“polarity-dependent cell elongation”
 - OP (Obsolete preferred term) [152,953 occurrences]
“Carbenoxolone sodium [gastro-intestinal use]”
“Acute abdomen”
 - OR (Orders) [1,357 occurrences]
All terms are from PCDS97 and are full utterances.

-
- “Discharge patient.”
- “Use assistive devices to maintain required position.”
- OSN (Official short name) [36,894 occurrences]
 - “Calculus Analysis”
 - “Drugs Ur Scn”
 - “Amoxicillin MIC”
 - PR (Name of a problem) [407 occurrences]
 - “Placenta abruptio”
 - “Dependency on alcohol”
 - PTGB (British preferred term) [21,376 occurrences]
 - “Abetalipoproteinaemia”
 - “Acting out - mental defence mechanism”
 - PX (Expanded preferred terms (pair with PS)) [3,157 occurrences]

These terms often contain characters indicating a superscript or subscript.

 - “Ca²⁺-transporting ATPase”
 - “alpha₁ Antichymotrypsin”
 - SBD (Semantic branded drug) [17,572 occurrences]
 - “Sodium Fluorescein 250 MG/ML Injectable Solution [AK-Fluor]”
 - “Acetaminophen 250 MG / Caffeine 30 MG / Chlorpheniramine 2 MG / Hydrocodone 5 MG / Phenylephrine 10 MG Oral Tablet [Hycomine Compound]”
 - SBDC (Semantic Branded Drug Component) [16,990 occurrences]
 - “aminodarone 200 MG [Sandoz brand of Amiodarone]”
 - “Fish Oils 1000 MG [Sea-Omega]”
 - SBDF (Semantic branded drug and form) [13,740 occurrences]
 - “Nitrofurazone Topical Cream [Furacin]”
 - “Dianeal Low Calcium with 1.5% Dextrose Intraperitoneal Solution”
 - SC (Special Category term) [53 occurrences]
 - “Congenital Malformations”
 - “bandages”
 - SCD (Semantic Clinical Drug) [30,276 occurrences]
 - “calcium chloride, dihydration 0.2 MG/ML / Potassium Chloride 0.3 MG/ML / Sodium Chloride 6 MG/ML / Sodium Lactate 3.1 MG/ML Irrigation Solution”
 - “Hydrogen Peroxide 300 MG/ML Topical Solution”
 - SCDC (Semantic Drug Component) [23,089 occurrences]
 - “Iron-Dextran Complex 100 MG/ML”
 - “ALLERGENIC EXTRACT, DUST, AUTOGENOUS 1 UNT/ML”
-

-
- SCDF (Semantic clinical drug and form) [14,288 occurrences]
 - “Triamcinolone Oral Paste”
 - “Ephedrine / Phenobarbital / Potassium Iodide / Theophylline Oral Tablet”
 - SD (CCS single-level diagnosis categories) [284 occurrences]
 - “Abdominal pain”
 - “Spontaneous abortion”
 - SI (Name of a sign or symptom of a problem) [311 occurrences]
 - “allergens”
 - “anemia”
 - SN (Official component synonym in LOINC) [319 occurrences]
 - Most of these terms are abbreviatory.
 - “ETOH”
 - “O NOS AG”
 - SP (CCS single-level procedure categories) [234 occurrences]
 - “Diagnostic amniocentesis”
 - “Abortion (termination of pregnancy)”
 - ST (Step) [132 occurrences]
 - “Manage contracts”
 - “Interview the patient or patient’s representative”
 - SX (Mixed-case component synonym with expanded abbreviations) [453 occurrences]
 - Like PX above, these terms contain special characters to indicate superscripts or subscripts.
 - 24,25-Dihydroxyvitamin D>3<
 - Vitamin B>4<
 - SYGB (British synonym) [7,634 occurrences]
 - “Tumour of abdomen”
 - “ABL - Abetalipoproteinaemia”
 - TA (Task) [170 occurrences]
 - “Introduce self to patient and explain services”
 - “Determine patient’s primary spoken language and communications ability/limitations”
 - TC (Term class) [61 occurrences]
 - “GI_NOS”
 - “ABDOMEN”
 - TG (Name of the target of an intervention) [63 occurrences]
 - “Behavior modification”
 - “Communication”

- TQ (Topical qualifier) [83 occurrences]
The meaning of these terms is specific to MeSH indexing and may not be appropriate for general use, but they are not currently being excluded because the Medical Text Indexer (MTI) includes them in its recommendations.
“abnormalities”
“administration & dosage”
- TX (CCPSS synthesized problems for TC termgroup) [61 occurrences]
“GI_NOS PROBLEM”
“ABDOMEN PROBLEM”
- UE (Unique equivalent name) [3 occurrences]
“Tanella <Bacteria>”
“Wernerella <Bacteria>”
- XD (Expanded descriptor in AOD) [509 occurrences]
“identification and screening for Alcohol or Other Drugs use”
“Alcohol or Other Drug Disorder”

7.3 Good Types

This section contains the remaining types, i.e., those most appropriate for text processing.

- AC (Activities) [10,764 occurrences]
“Monitor blood pressure”
“Control bleeding”
- AD (Adjective) [988 occurrences]
“Anorexic”
“Scarred”
- AS (Attribute type synonym) [25 occurrences]
“Precipitating factor”
“Px - Prescription”
- AT (Attribute type) [819 occurrences]
“Allergen”
“Association”
- BioCPT (BioCarta PT) [300 occurrences]
“Telomeres, Telomerase, Cellular Aging, and Immortality”
“Chromatin Remodeling by hSWI/SNF ATP-dependent Complexes”
- BN (Fully-specified drug brand name that can not be prescribed) [26,801 occurrences]
“Parlodel”
“Aminocaproic Acid”

-
- BSY (Broad synonym) [630 occurrences]
“p53-mediated DNA damage response”
“ER stress response”
 - CDCPT (Centers for Disease Control and Prevention PT)
“Polynesian“
“Egyptian“
 - CE (Entry “term” to a Supplementary Chemical “term”) [198,073 occurrences]
“2 bromolysergic acid diethylamide”
“7S RNA”
 - CL (Class) [14 occurrences]
“Managing the Practice”
“Ensuring Appropriate Pharmacotherapy”
 - CMN (Common name) [7,421 occurrences]
“acanthocephalans”
“alfalfa”
 - CN (LOINC official component name) [16,103 occurrences]
“DIPALMITOYLPHOSPHATIDYLCHOLINE”
“1,4-ALPHA GLUCAN BRANCHING ENZYME”
“???LEAD” (*sic*)
 - CU (Common usage) [10 occurrences]
“Entelev”
“Jim’s Juice”
 - CX (Component process in ICPC with abbreviations expanded) [5,763 occurrences]
“NOS ANTIBODY”
“CANCER ANTIGEN 125”
 - DCPPT (Division of Cancer Prevention Program PT) [607 occurrences]
“2-Mercaptoethanesulfonate, Sodium Salt“
“(17alpha)-Pregna-2,4-dien-20-yno[2,3-d]isoxazol-17-ol“
 - DE (Descriptor) [10,817 occurrences]
“synthetic 11-hydroxycorticosteroids”
“abdomen”
 - DF (Dose Form) [840 occurrences]
“24 Hour Transdermal Patch”
“Aerosol”

-
- DI (Disease name) [6,172 occurrences]
“ABETALIPOPROTEINEMIA”
“Abortion, Spontaneous”
 - “DN” (Display Name) [8 occurrences]
“B16 Malignant Melanoma”
“Chloroacetate Esterase Staining Method”
 - DO (Domain) [4 occurrences]
“Health Systems Management”
“Ensuring Appropriate Therapy and Outcomes”
 - 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) [177 occurrences]
“Acetylcholinesterase”
“Animal”
 - DTPSY (Developmental Therapeutics Program SY) [2,040 occurrences]
“Dimethylbenzanthracene”
“3-Hydroxy-L-tyrosine”
 - DX (Diagnosis) [256 occurrences]
“Anxiety”
“Blood Pressure Alteration”
 - EN (MeSH nonprint entry “term”) [31,154 occurrences]
“(131)I-MAA”
“Injuries, Abdominal”
 - EP (Entry “term”) [17,682 occurrences]
“Dipalmitoyllecithin”
“Branching Enzyme”
 - ESY (Exact synonym) [14,241 occurrences]
“ageing”
“senescence”
 - ET (Entry “term”) [45,431 occurrences]
“MPTP”
“Abelson’s virus”
 - EX (Expanded form of entry term) [20,895 occurrences]
“periodic light Alcohol or Other Drugs use”
“Alcohol or Other Drugs tax”

-
- FBD (Foreign brand name) [1,153 occurrences]
 - “Helixor”
 - “Iscador”
 - FDAAB (Food and Drug Administration AB) [438 occurrences]
 - “T-DERMAL”
 - “I-LESION”
 - FDAPT (Food and Drug Administration PT) [2,958 occurrences]
 - “MENSA”
 - “EPINEPHRINE”
 - FDASY (Food and Drug Administration SY) [627 occurrences]
 - “Replacement of aortic valve”
 - “Other (For Use When An Appropriate Patient Code Cannot Be Identified)”
 - FI (Finding name) [5,016 occurrences]
 - “ABDOMINAL PAIN, CRAMPY”
 - “ABDOMINAL DISTENTION”
 - GN (Generic drug name) [2,548 occurrences]
 - “mesna”
 - “paraaminobenzoic acid”
 - GT (Glossary “term”) [4,797 occurrences]
 - “SYNDROME ABDOMINAL ACUTE”
 - “ABDOMINAL CRAMP”
 - HC (Hierarchical class) [443 occurrences]
 - “Behavior Therapy”
 - “Cognition”
 - HG (High Level Group Term) [367 occurrences]
 - “Adrenal gland disorders”
 - “Benign neoplasms gastrointestinal”
 - HT (Hierarchical term) [22,554 occurrences]
 - “Abdomen”
 - “Abdominal pain”
 - ID (Nursing indicator) [2,576 occurrences]
 - “Ankylosed joints”
 - “Appetite loss”

-
- IN (Name for an ingredient) [26,811 occurrences]
 - “mesna”
 - “beta-Alanine”
 - INP (Ingredient preparation) [4,231 occurrences]
 - “MESNA PREPARATION”
 - “PARA-AMINO BENZOIC ACID PREPARATION”
 - IT (Index “term”, i.e., derived from the index to any non-MeSH source vocabulary) [2,077 occurrences]
 - “ACUTE ABDOMEN”
 - “CRAMP ABDOMINAL”
 - IV (Intervention) [849 occurrences]
 - “Activities of Daily Living (ADLs)”
 - “Pain Management”
 - KEGGPT (Kyoto Encyclopedia of Genes and Genomes PT) [125 occurrences]
 - “Alzheimer’s Disease”
 - “tRNA Biosynthesis (Eukaryotes)”
 - LPN (LOINC parts name) [16,598 occurrences]
 - “DIPALMITOYLPHOSPHATIDYLCHOLINE”
 - “1,4-ALPHA GLUCAN BRANCHING ENZYME”
 - LT (Lower Level Term) [55,621 occurrences]
 - “Acute abdomen”
 - “X-ray NOS abnormal”
 - LV (Lexical variant) [10 occurrences]
 - “Hodgkin lymphoma”
 - “Hodgkins lymphoma”
 - MD (CCS multi-level diagnosis categories) [693 occurrences]
 - “Abdominal pain”
 - “Congenital anomalies”
 - MH (Main heading) [24,357 occurrences]
 - “1,2-Dipalmitoylphosphatidylcholine”
 - “Abdomen”
 - MOA (Mechanism of action) [256 occurrences]
 - “Adrenergic alpha-Agonists”
 - “Bile Acids”

-
- MS (Multum names of branded and generic supplies or supplements) [7,207 occurrences]
“Acetone”
“0.3cc Syringe 29g 1/2””
 - MTH_BSY (Metathesaurus broad synonym expanded) [11 occurrences]
“2-oxoglutarate dehydrogenase complex location”
“CBC complex location”
 - MTH_ESY (Metathesaurus exact synonym expanded) [342 occurrences]
“Golgi complex location”
“Skp1/Cul1/F-box protein complex location”
 - MTH_HG (MTH High Level Group Term) [51 occurrences]
“Hemoglobinopathies”
“Leukemias”
 - MTH_HT (MTH Hierarchical term) [192 occurrences]
“Amebic infections”
“Carcinoid tumors”
 - MTH_LT (MTH Lower Level Term) [565 occurrences]
“Myelosis-non-leukemic”
“Alpha thalassemia”
 - MTH_NPT (MTH non-preferred for language term) [2 occurrences]
“Zuni”
“Diegueno”
 - MTH_NSY (Metathesaurus narrow synonym expanded) [106 occurrences]
“anaphase promoting complex location inhibition”
“succinate-CoA ligase complex location (GDP-forming) (sensu Bacteria)”
 - MTH_PT (MTH Designated preferred name) [4,597 occurrences]
“Anemia hemolytic autoimmune”
“Local anesthesia”
 - MTH_PTN (Metathesaurus preferred term, natural language form) [512 occurrences]
“hepatitis A & B immunization”
“injection of influenza immunization”
 - MTH_RLS (Metathesaurus related synonym expanded) [30 occurrences]
“distal pole complex location”
“single-stranded DNA-binding protein complex location”

-
- N1 (Chemical Abstracts Service Type 1 name of a chemical) [22,546 occurrences]
“1,4-alpha-D-Glucan:1,4-alpha-D-glucan 6-alpha-D-(1,4-alpha-D-glucano)-transferase”
“1,1,3-Propanetricarboxylic acid, 3-amino-”
 - NCI-GLOSSPT (NCI-GLOSS PT) [2,653 occurrences]
“mercaptopurine”
“pernicious anemia”
 - NCI-GLOSSSY (NCI-GLOSS SY) [53 occurrences]
“cell differentiation“
“IL-1-alfa“
 - NM (Supplementary chemical “term”, a name of a substance) [165,179 occurrences]
“2-bromolysergic acid diethylamide”
“3-hydroxyproline”
 - NP (Non-preferred term) [5,437 occurrences]
“3,4-methylenedioxyamphetamine”
“congenital defects”
 - NSY (Narrow synonym) [1,736 occurrences]
“surface antigen variation”
“bone calcification”
 - OBS (Obsolete broad synonym) [12 occurrences]
“general amino acid transporter”
“oligomerization activity”
 - OPN (Obsolete preferred term, natural language form) [85 occurrences]
“carcinoma of the large intestine”
“acquired pylorospasm“
 - ORS (Obsolete related synonym) [26 occurrences]
“COPI vesicle”
“FK506 binding protein”
 - OS (System-organ class in the WHO Adverse Reaction Terminology) [58 occurrences]
“PSYCHIATRIC DISORDERS”
“AUTONOMIC NERVOUS SYSTEM DISORDERS”
 - PC (Preferred “trimmed term” in ICPC) [233 occurrences]
“arthrogryposis multiplex congenita”
“Bartter syndrome”

-
- PCE (Preferred entry “term” to a Supplementary Concept “term”) [62,518 occurrences]
“Adona”
“aquocobalamin acetate”
 - PE (Physiologic effect) [1,699 occurrences]
“Vasoconstriction”
“Cardiac Rate Alteration”
 - PEN (Preferred MeSH nonprint entry “term”) [14,660 occurrences]
“(131)I-Macroaggregated Albumin”
“Abortion Centers”
 - PEP (Preferred entry “term”) [6,994 occurrences]
“17 beta-Hydroxysteroid Dehydrogenases”
“Abdominal Cramps”
 - PK (Pharmacokinetics) [59 occurrences]
“Absorption”
“Glutathione s-transferases”
 - PM (Machine permutation) [81,456 occurrences]
“1,2 Dipalmitoylphosphatidylcholine”
“Enzyme, Branching”
 - PN (Metathesaurus preferred name) [85,600 occurrences]
“17-Hydroxysteroid Dehydrogenases”
“Droxidopa”
 - PQ (Qualifier for a problem) [9 occurrences]
“Family”
“Health Promotion”
 - PSC (Protocol selection criteria) [555 occurrences]
“bladder cancer”
“basal cell carcinoma of the skin”
 - PT (Designated preferred name) [1,004,075 occurrences]
“Dipalmitoylphosphatidylcholine”
“Abdomen”
 - PTN (Preferred term, natural language form) [7,410 occurrences]
“keratoconus”
“solar keratosis”

- PXQ (Preferred term in preferred qualifier concept) [168 occurrences]
“agenesis”
“wounds”
- RLS (Related synonym) [1,128 occurrences]
“blood coagulation factor activity”
“glycine betaine/proline porter activity”
- RS (Extracted related names in SNOMED2) [68 occurrences]
“Aleutian disease virus”
“Aluminum silicate”
- RT (Designated related “term”) [6,821 occurrences]
“Lumpy jaw”
“20-Hydroxyprogesterone”
- SCN (Scientific name) [245,487 occurrences]
“Abelson murine leukemia virus”
“Absidia”
- SS (Synonymous “short” forms) [196 occurrences]
“Alzheimer disease”
“adrenoleukodystrophy”
- SY (Designated synonym) [400,328 occurrences]
“Branching enzyme”
“Amylo-(1,4,6)-transglycosylase”
- XQ (Alternate name for a qualifier) [67 occurrences]
These terms are similar to TQ terms, and the same comments apply.
“anomalies”
“teratology”