

Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus

J.J. Cimino,^{a,*} H. Min,^b and Y. Perl^b

^a Department of Medical Informatics, Columbia University, New York, NY, USA

^b Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA

Received 21 October 2002

Abstract

Objective. To develop and test a method for automatically detecting inconsistencies between the parent–child is-a relationships in the Metathesaurus and the ancestor–descendant relationships in the Semantic Network of the Unified Medical Language System (UMLS).

Methods. We exploited the fact that each Metathesaurus concept is assigned one or more semantic types from the UMLS Semantic Network and that the semantic types are arranged in a hierarchy. We compared the semantic types of each pair of parent and child concepts to determine if the types “explained” the Metathesaurus is-a relationships. We considered cases where the semantic type of the parent was neither the same as, nor an ancestor of, the semantic type of the child to be “unexplained.” We applied this method to the January 2002 release of the UMLS and examined the unexplained cases we discovered to determine their causes.

Results. We found that 17,022 (24.3%) of the parent–child is-a relationships in the UMLS Metathesaurus could not be explained based on the semantic types of the concepts. Causes for these discrepancies included cases where the parent or child was missing a semantic type, cases where the semantic type of the child was too general or the semantic type of the parent was too specific, cases where the parent–child relationship was incorrect, and cases where an ancestor–descendant relationship should be added to the UMLS Semantic network. In many cases, the specific cause of the discrepancy cannot be resolved without authoritative judgment by the UMLS developers.

Conclusions. Our method successfully detects inconsistencies between the hierarchies of the UMLS Metathesaurus and Semantic Network. We believe that our method should be added to the set of tools that the UMLS developers use to maintain and audit the UMLS knowledge sources.

© 2003 Elsevier Inc. All rights reserved.

1. Introduction

The Unified Medical Language System (UMLS) [1,2], developed by the National Library of Medicine (NLM), includes two knowledge resources, the Metathesaurus (Meta) [3,4] and the Semantic Network (SN) [5,6], that comprise a complex knowledge base of medical concepts drawn from over 100 terminologies. Each of these resources includes hierarchical information: the SN organizes semantic types in a strict is-a hierarchy, while Meta utilizes a variety of hierarchical relationships between pairs of concepts. The two resources are connected by the assignment of one or more semantic types

from the SN to each concept in Meta. While the creation and maintenance of these resources, their structures, and interrelationships is daunting (Meta contains approximately 800,000 concepts), automated tools are employed to assist human reviewers with the management tasks [7].

The management of the UMLS content is of the utmost importance to its users, who depend on its quality for performance of their systems. For example, a popular use of the UMLS is to support searching bibliographic databases. One search method that can exploit UMLS knowledge involves the “explosion” of a general term into an OR’ed list of its descendant terms. If a literature search is done by exploding a term that has incorrect terms under it in the terminology’s hierarchy, inappropriate terms will be included in the search

* Corresponding author. Fax: 1-212-305-3302.

E-mail address: jjc7@columbia.edu (J.J. Cimino).

strategy with a corresponding reduction in the relevance of retrieved results.

For this reason, and many others, the UMLS developers strive for accuracy in the UMLS knowledge sources. They are aided in their task by the knowledge in Meta and the SN, which can be leveraged to support their maintenance through automated auditing for internal consistency [8–16]. In some cases, the auditing can pinpoint inconsistencies and outright errors that can be readily addressed. In other cases, the methods used can merely suggest potential problems. In all cases, however, the automated methods can help to focus the limited resources of human review to the cases most likely to need attention.

In this paper, we describe an approach that compares the parent–child relationships between concepts in Meta with the ancestor–descendant¹ relationships between semantic types in the SN to identify inconsistencies in Meta and suggest changes to the SN. We report the results of applying this method to the 2002 UMLS and discuss their implications.

2. Background

The basic unit of information in Meta is the concept, which is identified by a Concept Unique Identifier (CUI). When terms from disparate terminologies are found to be synonymous, they are merged into single concepts. Each concept in Meta is then assigned one or more semantic types from the SN. For example, the concept **Organ** (C0178784)² can be found in four different UMLS source terminologies and has been assigned the semantic type *Body Part, Organ, or Organ Component* (T023, A.1.2.3.1). A second concept, **Anatomic structures** (C0700276), comes from five different terminologies and has the semantic type *Anatomical Structure* (T017, A1.2).

Meta includes a variety of relationships between concepts, provided in a file called MRREL. Relationships include **parent–child**, **broader–narrower**, **like**, and **other**; they may be further characterized with specific semantic relationships, such as **is-a** and **part-of**. For example, MRREL includes a **parent–child is-a** relationship between **Organ** and **Anatomic Structures**, indicating that the former is a more specific concept than the latter.

¹ To avoid confusion, we will refer to the hierarchical relationships in Meta as “is-a” and those in the SN as “ancestor–descendant.”

² In this paper, Meta concepts and relations will be depicted in **bold**; concepts have Concept Unique Identifiers (CUIs) composed of a “C” and seven numeric digits. Semantic types and relations from the SN will be depicted in *italics*; semantic types are identified with a code composed of a “T” and three numeric digits and with a tree address composed of a letter followed by one or more digits separated by periods.

The semantic types in the SN are arranged in a strict hierarchy (that is, each may have at most one parent) of *ancestor–descendant* relationships, implicit in the tree addresses provided for each semantic type. For example, in the current SN, *Anatomical Structure* (with tree address A1.2) is the immediate *ancestor-of Fully Formed Anatomical Structure* (with tree address A.1.2.3) that, in turn, is the immediate *ancestor-of Body Part, Organ, or Organ Component* (with tree address A.1.2.3.1).

A number of researchers have exploited UMLS knowledge to help with auditing Meta. For example, Gu et al. [10] and Bodenreider [12] have reported the use of UMLS knowledge to construct object-oriented models that support navigation of Meta, with potential usefulness for maintenance. In a third study, Bodenreider used the hierarchical information in Meta to detect and remove circular relationships [9]. In a fourth study, McCray and Bodenreider [16] compared the associative and hierarchical relationships between concepts in a subset of Meta with the allowable relationships based on the Semantic Net. Finally, one of us has used the coupling of Meta concepts and semantic types to detect ambiguity, inconsistent **parent–child** relationships and additional semantic relations for the SN [8,13].

3. Methods

The presence of hierarchies in both the SN and Meta, and the tight connection between the semantic types and the concepts, suggests a certain symmetry. Given the meaning of “is a” (both in plain English and in formal knowledge representation), if **Concept 1 is-a Concept 2** it seems reasonable to assume that both concepts are either of the same semantic type, or else the type of **Concept 1** should have an *ancestor–descendant* relationship to the type of **Concept 2**, either immediate or indirect.³ Indeed, this is the case with the example presented above: **Organ is-a Anatomic Structures**, and *Body Part, Organ, or Organ Component is-descendant-of Anatomical Structure*.

Our approach examines cases where there is an inconsistency between the semantic types assigned to concepts in Meta that have an **is-a** relationship. Specifically, we look at all instances in MRREL where none of the semantic types of the parent concept is identical to, or an *ancestor-of*, any of the semantic types of the child concept. Fig. 1 shows examples of “expected relationships” between concept pairs, based on semantic types, and Fig. 2 shows examples of “unexpected relationships” between concept pairs, based on semantic types.

³ The *ancestor–descendant* relationship is transitive; since *Body Part, Organ, or Organ Component is-descendant-of Fully Formed Anatomical Structure*, and *Fully Formed Anatomical Structure is-descendant-of Anatomical Structure*, it is also implied that *Body Part, Organ, or Organ Component is-descendant-of Anatomical Structure*.

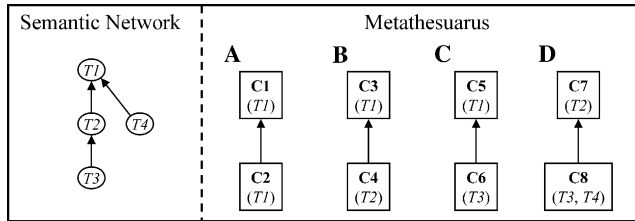


Fig. 1. Examples of expected *is-a* relationships between concepts. A hypothetical subtree of the UMLS Semantic Network is shown on the left, consisting of four semantic types (*T1–T4*). A hypothetical piece of the UMLS Metathesaurus is shown on the right, consisting of eight concepts (*C1–C8*), arranged in pairs of *is-a* relationships (shown by arrows). Each concept is assigned one or more semantic types, shown in parentheses. In **A**, both concepts have the same semantic type. In **B** and **C**, the relationship between the two semantic types is *ancestor–descendant*. In **D**, the child concept has two semantic types, but since one of them (*T3*) is a *descendant-of* the parent concept's type (*T2*), this is an expected relationship.

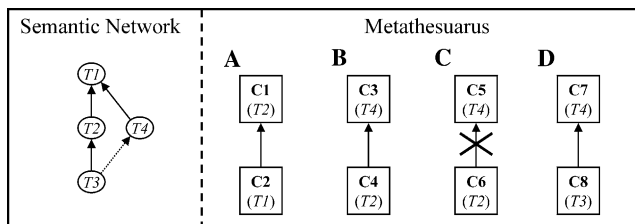


Fig. 2. Examples of unexpected *is-a* relationships between concepts. In **A**, the semantic type (*T2*) of the parent concept is a *descendant-of* the semantic type (*T1*) of the child concept, suggesting that either the parent concept's type is too specific or the child concept's type is too general. In **B**, the semantic types of the two concepts are neither the same nor in an *ancestor–descendant* relationship, suggesting that one or the other concepts is missing a type (i.e., *C3* might be missing *T1* or *T2*, or *C4* might be missing *T4*). **C** also has unrelated semantic types in the two concepts, but in this case the explanation is that the *is-a* relationship between *C5* and *C6* is incorrect. **D** also has unrelated semantic types in the two concepts, but in this case the explanation is that *T4* is conceptually an *ancestor-of* *T3* (shown by the dotted arrow) but is not included in the UMLS Semantic Network.

We consider that each unexpected relationship can be explained by one or more of six causes:

- (1) *Parent-Too-Specific*: the semantic type of the parent concept is a *descendant-of* the semantic type of the child concept; if the parent concept was assigned a less specific semantic type, the *is-a* relationships between the concepts would be expected (see Fig. 2A).
- (2) *Child-Too-General*: the semantic type of the child concept is an *ancestor-of* the semantic type of the parent concept; if the child concept was assigned a more specific semantic type, the *is-a* relationship between the concepts would be expected (see Fig. 2A).
- (3) *Parent-Type-Missing*: if the parent concept was assigned an additional semantic type, the *is-a* relationship between the concepts would be expected (see Fig. 2B).

- (4) *Child-Type-Missing*: if the child concept was assigned an additional semantic type, the *is-a* relationship between the concepts would be expected (see Fig. 2B).
- (5) *Wrong-Is-A*: the *is-a* relationship between the concepts is incorrect (see Fig. 2C).
- (6) *Missing-Ancestor–Descendant*: if an *ancestor–descendant* link was added to the Semantic Network, the *is-a* relationship between the concepts would be expected (see Fig. 2D).

While automated methods can be used to detect inconsistencies, automatic determination of the specific reason for each case is not generally possible. For example, if the semantic type of the child concept is an *ancestor-of* the semantic type of the parent concept, there is no way to automatically determine whether the problem is *Parent-Too-Specific* or *Child-Too-General* without human review. This review, in turn, depends on the definitions of the semantic types and (where available) the definitions of the concepts.

To conduct our review, we extracted all the records in MRREL in which the relationship was “CHD” (a **child-of** relationship) and the relationship attribute was “is-a.” These records contain two CUIs, CUI1 and CUI2, for which the relationship is **CUI1 is-a CUI2**. We obtained the preferred English name for each CUI from the file MRCON.

We obtained all semantic types associated with each of the CUIs from the file MRSTY and aggregated the concept pairs into “relationship sets” based on the semantic types of the parent and child concepts. Relationships involving concepts with multiple semantic types were aggregated into multiple relationship sets. We then obtained the names and tree addresses of each semantic type from the file SRDEF. Details of MRREL, MRCON, MRSTY, and SRDEF can be found in the UMLS documentation [2].

Once we obtained the relationship sets, we identified those that represented expected relationships. These were cases where the semantic type of the parent concepts was either identical to, or an *ancestor-of*, the semantic type of the child concepts. We determined this by examining the tree addresses. For example the tree address for *Entity* (T071) is “A,” and the tree address for *Intellectual Product* (T170) is “A2.4.” Since “A2.4” has the prefix “A,” we can infer that *Intellectual Product is-descendant-of Entity* in the SN; therefore, the set of relationships from MRREL in which the parent concepts have the type *Entity* and the child concepts have the type *Intellectual Product* is expected. Conversely, since “A” has no prefix “A2.4,” the set of relationships from MRREL in which the parent concepts have the type *Intellectual Product* and the child concepts have the type *Entity* is not expected (see Fig. 2A). We manually examined the unexpected relationship sets to try to understand why they were

occurring (that is, which of the six causes listed above were present).

4. Results

We used the January 2002 release of the UMLS. Of the 10,147,419 records in MRREL, 654,292 had the relationship “CHD”; of these, 69,991 had the **is-a** relationship attribute. These records involved 20,442 unique parent codes and 67,453 unique children codes, with 67,589 unique codes overall (since most parent concepts were also children). These concepts had a total of 68,192 semantic types in MRSTY. After merging concept pairs into relationship sets based on their semantic types and excluding expected relationship sets, there remained 17,022 relationships in 246 relationship sets. The largest relationship sets, containing over 30 concept pairs, are shown in Table 1. These 34 relationship sets represent

13.8% of the 246 relationship sets and account for 16,256 (95.5%) of the 17,022 concept pairs.

4.1. Clinical Drug relationship sets

The largest unexplained relationship set involves parent concepts of type *Pharmacologic Substance* (T121, A1.4.1.1.1) and child concepts of type *Clinical Drug* (T200, A1.3.3); this one set contains 9296 occurrences, accounting for 54.6% of the unexplained relationships. Fig. 3 depicts one example, **Antifungal Agents** (C0003308) and its child **FLUCONAZOLE 100 MG ORAL TABLET** (C0688874). *Clinical Drug* is defined as “A pharmaceutical preparation as produced by the manufacturer” and is an immediate *descendant-of Manufactured Object* (T073, A1.3) in the SN. We therefore believe that each of the members of this relationship set is an example of the presence of a Wrong-Is-A in MRREL.

Table 1
Unexplained relationship sets with greater than 30 concept pairs

Semantic Type of Parent Concept		Semantic Type of Children Concepts		Pairs	Cause	
T121	A1.4.1.1.1	Pharmacologic Substance	T200 A1.3.3	Clinical Drug	9296	Wrong-Is-A
T109	A1.4.1.2.1	Organic Chemical	T200 A1.3.3	Clinical Drug	723	Wrong-Is-A
T029	A2.1.5.2	Body Location or Region	T023 A1.2.3.1	Body Part, Organ, or Organ Component	721	Missing-Type
T127	A1.4.1.1.3.4	Vitamin	T200 A1.3.3	Clinical Drug	582	Wrong-Is-A
T121	A1.4.1.1.1	Pharmacologic Substance	T074 A1.3.1	Medical Device	412	Wrong-Is-A
T195	A1.4.1.1.1.1	Antibiotic	T200 A1.3.3	Clinical Drug	362	Wrong-Is-A
T125	A1.4.1.1.3.2	Hormone	T200 A1.3.3	Clinical Drug	331	Wrong-Is-A
T130	A1.4.1.1.4	Indicator, Reagent, or Diagnostic Aid	T200 A1.3.3	Clinical Drug	286	Wrong-Is-A
T104	A1.4.1.2	Chemical Viewed Structurally	T200 A1.3.3	Clinical Drug	281	Wrong-Is-A
T030	A2.1.5.1	Body Space or Junction	T029 A2.1.5.2	Body Location or Region	261	Missing-Type
T023	A1.2.3.1	Body Part, Organ, or Organ Component	T029 A2.1.5.2	Body Location or Region	230	Missing-Type
T029	A2.1.5.2	Body Location or Region	T030 A2.1.5.1	Body Space or Junction	228	Missing-Type
T122	A1.4.1.1.2	Biomedical or Dental Material	T200 A1.3.3	Clinical Drug	226	Wrong-Is-A
T123	A1.4.1.1.3	Biologically Active Substance	T200 A1.3.3	Clinical Drug	219	Wrong-Is-A
T196	A1.4.1.2.3	Element, Ion, or Isotope	T200 A1.3.3	Clinical Drug	219	Wrong-Is-A
T110	A1.4.1.2.1.9.1	Steroid	T200 A1.3.3	Clinical Drug	197	Wrong-Is-A
T082	A2.1.5	Spatial Concept	T023 A1.2.3.1	Body Part, Organ, or Organ Component	178	Missing-Type
T024	A1.2.3.2	Tissue	T023 A1.2.3.1	Body Part, Organ, or Organ Component	174	Missing-Type
T116	A1.4.1.2.1.7	Amino Acid, Peptide, or Protein	T200 A1.3.3	Clinical Drug	158	Wrong-Is-A
T197	A1.4.1.2.2	Inorganic Chemical	T200 A1.3.3	Clinical Drug	153	Wrong-Is-A
T118	A1.4.1.2.1.8	Carbohydrate	T200 A1.3.3	Clinical Drug	148	Wrong-Is-A
T129	A1.4.1.1.3.5	Immunologic Factor	T200 A1.3.3	Clinical Drug	131	Wrong-Is-A
T130	A1.4.1.1.4	Indicator, Reagent, or Diagnostic Aid	T074 A1.3.1	Medical Device	122	Wrong-Is-A
T023	A1.2.3.1	Body Part, Organ, or Organ Component	T024 A1.2.3.2	Tissue	102	Missing-Type
T030	A2.1.5.1	Body Space or Junction	T023 A1.2.3.1	Body Part, Organ, or Organ Component	96	Missing-Type
T022	A2.1.4.1	Body System	T023 A1.2.3.1	Body Part, Organ, or Organ Component	86	Missing-Type
T023	A1.2.3.1	Body Part, Organ, or Organ Component	T030 A2.1.5.1	Body Space or Junction	68	Missing-Type
T120	A1.4.1.1	Chemical Viewed Functionally	T200 A1.3.3	Clinical Drug	51	Wrong-Is-A
T077	A2	Conceptual Entity	T023 A1.2.3.1	Body Part, Organ, or Organ Component	43	Missing-Type
T129	A1.4.1.1.3.5	Immunologic Factor	T074 A1.3.1	Medical Device	43	Wrong-Is-A
T047	B2.2.1.2.1	Disease or Syndrome	T046 B2.2.1.2	Pathologic Function	33	Child-Too-General
T116	A1.4.1.2.1.7	Amino Acid, Peptide, or Protein	T074 A1.3.1	Medical Device	33	Wrong-Is-A
T023	A1.2.3.1	Body Part, Organ, or Organ Component	T022 A2.1.4.1	Body System	32	Missing-Type
T125	A1.4.1.1.3.2	Hormone	T074 A1.3.1	Medical Device	31	Wrong-Is-A

The “Cause” column identifies the major reason(s) which we believe explain each relationship set; “Missing-Type” includes Parent-Missing-Type and Child-Missing-Type.

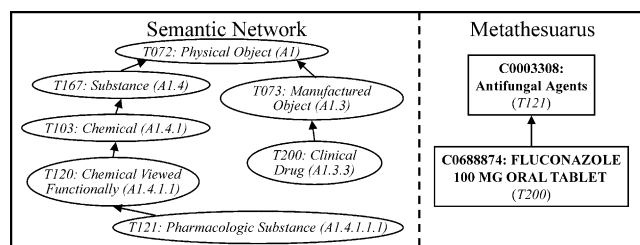


Fig. 3. One example of an unexplained relationship found in the UMLS Metathesaurus. The parent concept, **Antifungal Agents**, has semantic type *Pharmacologic Substance*, while the child concept, **FLUCONAZOLE 100 MG ORAL TABLET**, has semantic type *Clinical Drug*. There is currently no ancestor-descendant relationship between these two semantic types, as shown in the hierarchy at left. See the text for a discussion of possible explanations.

The *Clinical Drug/Pharmacologic Substance* relationship set is the largest of 20 relationship sets in which the child concepts are of semantic type *Clinical Drug* and the parent concepts have semantic types that are *descendant-of Chemical* (T103, A1.4.1). The 19 other unexplained relationship sets (15 of which have over 30 concept pairs and are shown in Table 1), involve an additional 4123 concept pairs; we believe these sets also represent cases of *Wrong-Is-A*.

An alternative possible explanation for these relationship sets is *Parent-Type-Missing*; correcting this cause would require assigning the semantic type *Clinical Drug* to concepts such as **Antifungal Agents**. Another possible cause is *Child-Type-Missing*; correcting this cause would require assigning some semantic type from the *Chemical* subtree of the SN to concepts such as **FLUCONAZOLE 100 MG ORAL TABLET**. The third possibility is *Missing-Ancestor-Descendant*; correction would require adding a *descendant-of* relationship between *Clinical Drug* and 20 different *descendants-of Chemical*. We believe that each of these solutions would be a violation of the UMLS's definition of *Clinical Drug*. The information in Meta supports this view, since the majority of the 81,165 *Clinical Drug* concepts in Meta are not involved in these unexplained relationships.

4.2. Medical Device relationship sets

Like *Clinical Drug*, the semantic type *Medical Device* (T074, A1.3.1) is an immediate *descendant-of Manufactured Object*. As with *Clinical Drug*, many concepts with the semantic type *Medical Device* have parent concepts that have a semantic type in the *Chemical* subtree of the SN. We found 667 such concept pairs that were contained in 11 relationship sets (five sets have over 30 concept pairs and are shown in Table 1). We believe that these, too, represent cases of *Wrong-Is-A*.

4.3. Body Part, Organ or Organ Component relationship sets

There are 14 unexplained relationship sets (four shown in Table 1), containing 485 concept pairs, in which the parent concepts have the semantic type *Body Part, Organ or Organ Component* (T023, A1.2.3.1). An additional 11 unexplained relationship sets (six shown in Table 1), containing 1336 concept pairs, have child concepts with the semantic type *Body Part, Organ or Organ Component*. We believe that most of the unexplained concept pairs are cases of *Parent-Type-Missing* or *Child-Type-Missing*; our review of these 25 relationship sets support this view.

For example, **Capillary Bed** (C0489802) has the semantic type *Body Part, Organ, or Organ Component* and is the parent of **Systemic Capillary Bed** (C0923301), with semantic type *Body System*. It is our judgment that **Systemic Capillary Bed** should also have the semantic type *Body Part, Organ, or Organ Component* (*Child-Type-Missing*).

In another example, **Cardiac venous tree** (C0923573) has the semantic type *Body System* (T022, A2.1.4.1) and is the parent of **Smallest cardiac veins** (C0226663), with semantic type *Body Part, Organ, or Organ Component* (T023, A1.2.3.1). It is our judgment that **Cardiac venous tree** should also have the semantic type *Body Part, Organ, or Organ Component* (*Parent-Type-Missing*).

There are some cases where the *is-a* relationship between concepts appears to be wrong. For example, **Skeletal System of Upper Limb** (C081854), a *Body System*, is listed as a parent of **Bony pelvic girdle** (C0934859), a *Body Part, Organ, or Organ Component*. We judge that no changes of semantic type assignments will make **Bony pelvic girdle is-a Skeletal System of Upper Limb** a correct *is-a* relationship (*Wrong-Is-A*).

Two of the relationship sets in which the parent concepts have semantic type *Body Part, Organ, or Organ Component* (T023, A1.2.3.1) are special cases. One set has 22 concept pairs in which the child concepts have the semantic type *Fully Formed Anatomical Structure* (T021, A1.2.3); an example is **Right big toe** (C0930961) *is-a* **Hallux** (C0018534). The other set has one concept pair in which the child concept has the semantic type *Anatomical Structure* (T017, A1.2): **External rectal venous plexus** (C0580083) *is-a* **Rectal venous plexus** (C0580081). Because the tree address of *Body Part, Organ, or Organ Component* has as prefix the tree addresses of the other two semantic types, it is a descendant of *Fully Formed Anatomical Structure* and *Anatomical Structure*, similar to Fig. 2A. We believe that both these sets can be resolved by changing the semantic type of the children (e.g., **Right big toe** and **External rectal venous plexus**) from *Fully Formed Anatomical Structure* to *Body Part, Organ, or Organ Component*—cases of *Child-Too-General*.

4.4. *Body Location or Region and Body Space or Junction relationship sets*

One relationship set has 228 concept pairs in which the semantic type of the parent concepts is *Body Location or Region* (T029, A2.1.5.2) and the semantic type of the child concepts is *Body Space or Junction* (T030, A2.1.5.1). A second relationship set has 261 concept pairs that have the opposite semantic type assignments. For example, **Right inguinal canal** (C0459928), with semantic type *Body Space or Junction*, **is-a Inguinal Canal** (C0021445), with semantic type *Body Location or Region*. Conversely, **Middle ethmoidal cell** (C0928857), with semantic type *Body Location or Region*, **is-a Sinus** (C0030471), with semantic type *Body Space or Junction*. We believe that the concepts in these two sets should each have both semantic types (Parent-Type-Missing and Child-Type-Missing).

4.5. *Disease or Syndrome and Pathologic Function relationship set*

The previous four categories account for 33 of the 34 large relationship sets shown in Table 1. The remaining relationship set contains 33 concept pairs in which the parent concepts have semantic type *Disease or Syndrome* (T047, B2.2.1.2.1) and the child concepts have the semantic type *Pathologic Function* (T046, B2.2.1.2). For example, **Infertility, Male** (C0021364) **is-a Infertility** (C0021359). We believe that **Infertility, Male** and the other 32 children concepts in the set should have their semantic types changed from *Pathologic Function* to *Disease or Syndrome*—cases of Child-Too-General.

4.6. *Small unexplained relationship sets*

The above five categories cover the 34 relationship sets in Table 1 and 25 additional relationship sets (24.0% of the unexplained 246 relationship sets). Together, these sets cover 16,429 (96.5%) of the concept pairs. The remaining 593 concept pairs are grouped into 187 relationship sets. Table 2 shows the results of our analysis of 100 randomly selected concept pairs from this remaining group.

One systematic way to evaluate these sets is to identify those in which the semantic type of the parent

concepts *is-descendant-of* the semantic type of the child concepts (as was done for the *Body Part, Organ, or Organ Component Fully Formed Anatomical Structure* and *Disease or Syndrome/Pathologic Function* relationship sets described above) to determine if the cause is Parent-Too-Specific or Child-Too-General. Eighteen of the remaining relationship sets, containing 95 concept pairs, meet this criterion. We judged 12 of the relationship sets, containing 63 concept pairs, to be caused by Child-Too-General; for example, all 37 children concepts with semantic type *Spatial Concept* (T082, A2.1.5) should have the semantic type of their parent concepts (*Body Location or Region* (T029, A2.1.5.2) in 29 cases and *Spatial Concept* (T082, A2.1.5) in eight cases).

The remaining six of the above 18 relationship sets, containing 22 concepts, along with a random sample of the final 169 small relationship sets (summarized in Table 2), containing 498 concept pairs were due to a variety of causes, including Parent-Too-Specific, Parent-Type-Missing, Child-Type-Missing, and Wrong-Is-A. Specific counts of each cause are difficult to produce, however. Ambiguity in the meaning of the semantic types and concepts, as well as the intent of *is-a* and **is-a** relationships all contribute to this difficulty. Take, for example, the **is-a** relationship between **Arteriovenous Malformation** (C003857), with semantic type *Congenital Abnormality* (T019, A1.2.2.1) and its child term **Arteriovenous Fistula** (C0003855), with semantic type *Anatomical Abnormality* (T190, A1.2.2). Certainly arteriovenous fistulae are malformations of the arteriovenous system; some of them are congenital, but others are not, such as those that are created surgically [17]. But the term “arteriovenous malformation” is also used to refer to a very specific congenital abnormality. So, before the cause of this unexplained relationship can be resolved we need to know which meaning of “arteriovenous malformation” is intended. If both meanings are intended, then the ambiguous concept should be split into two concepts, for example **Congenital Arteriovenous Malformation** and **Congenital or Acquired Arteriovenous Malformation**. The former would have an **is-a** relationship to the latter, and the original **is-a** relationship would be preserved as **Arteriovenous Fistula is-a Congenital or Acquired Arteriovenous Malformation**.

4.7. *Missing-Ancestor–Descendant*

The structure of the SN is a particular interest of ours [18,19]. In fact, we undertook this study in part to seek evidence that the **is-a** relationships in Meta might support the addition or deletion of *ancestor–descendant* relationships in the SN. In our review of the results in this study, we succeeded in finding several relationship sets that seems to be due to the cause Missing-Ancestor–Descendant. The largest of these, with nine concept

Table 2
Reasons for unexplained relationship sets from a sample of small sets

Cause	Number of relationships
Child-Missing-Type	66
Parent-Missing-Type	18
Wrong-Is-A	6
Missing-Ancestor–Descendant	4
Child-Too-General	4
Parent-Too-Specific	2

pairs, has children concepts with semantic type *Injury or Poisoning* (T037, B2.3) and parent concepts with semantic type *Disease or Syndrome* (T047, B2.2.1.2.1). One example pair is **Inert Gas Narcosis is-a Occupational Disease**. We believe that the semantic types of both concepts are correct as is and that the **is-a** relationship between them is also correct. The only remaining explanation, then, is the inference that *Injury or Poisoning is-descendant-of Disease or Syndrome* should be added to the SN (Missing-Ancestor–Descendant). This set of nine concept pairs may seem to be scant supporting evidence; however, an additional 2186 **parent–child** relationships between concepts of these types can be found in MRREL. Although the relationship type is null, many of these may represent additional **is-a** pairs if the relationships types were to be made explicit. As a result, we have suggested to the NLM that they consider the addition of *Injury or Poisoning is-descendant-of Disease or Syndrome* to the SN.

5. Discussion

The majority of problems uncovered by our method were incorrect **is-a** relationships in the Meta hierarchy. Correction of such hierarchical errors is an important part of the UMLS maintenance, since many users rely on this knowledge for classification purposes. To take the example from Section 1, above, a user who wishes to search for articles about diseases of the **Skeletal System of Upper Limb**, and uses Meta to help with an “explode” function, may retrieve articles discussing the **Bony pelvic girdle**.

The addition of missing semantic type assignments, as well as removal of incorrect assignments, is also of great importance to UMLS users, who depend on such information for understanding how concepts from disparate terminologies are integrated in the Metathesaurus. Consider, for example, a case in which a UMLS user is constructing a list of prostheses terms listed in Meta. Since there is no semantic type “Prosthesis” in the SN, such concepts are appropriately categorized with the semantic types *Medical Device* and *Body Part, Organ and Organ Component*. Thus, a query of Meta for concepts with both semantic types will miss terms such as **Heart, Artificial**.

The method described in this paper is intended to provide a way for the UMLS developers to identify quickly one kind of inconsistency in their knowledge sources. We show that 24.3% of the relationships we examined in MRREL are unexplained. However, since we restricted our analysis to **parent–child is-a** relationships, this represents only 2.6% of all parent–child (“PAR”/“CHD”) relationships and about 0.3% of all the relationships in MRREL.⁴ The application of our

method to other kinds of relationships in MRREL will depend upon clarification of the semantics represented by the relationships. The vast majority (555,594 or 84.9%) of the PAR-CHD relationships are not further specified. If one assumes that the default **parent–child** relationship is also **is-a**, then our method could be extended to cover a much larger proportion (about 13%) of MRREL.

Our method is automated insofar as it identifies unexplained relationships, but it then requires manual review to identify the specific cause for each instance. The results of our manual review suggest several ways in which the method could be extended to further the automated process and reduce the burden of manual review. For example, by knowing that the semantic type of a child concept is the *ancestor-of* the semantic type of the parent concept (as in Fig. 2A), likely causes can be narrowed down to Parent-Too-Specific and Child-Too-General. Since the semantic typing in the UMLS is supposed to be as specific as possible [20], the most likely solution in each case will probably be to simply replace the semantic type of the child concept with the (more specific) type of the parent concept. Manual review then only needs to be done to confirm the appropriateness of each type assignment to the child concepts, as we have shown in Section 4. This reduces to four the number of possible causes for the remaining relationships.

Another way to simplify the review of unexplained relationships is to examine the relationship sets to determine if they are evidence for Missing-Ancestor–Descendant relationships in the SN. This requires analysis of only the semantic type pairs, not the concept pairs, in the relationship sets. In those sets where an *ancestor–descendant* relationship is missing from the SN, its addition will provide an explanation for all members of the set. In the remaining cases, the possible causes will now be reduced to three.

There may also be a way to automate the detection of Wrong-Is-A. Previous work has shown that some semantic types are mutually exclusive [13]. By considering this restriction, users of our method can automatically tell when no amount of addition of semantic types to the parent or child concepts will result in a correct **is-a** between them. For example, if we consider that *Clinical Drug* concepts are manufactured objects, then such concepts could never be classified as any semantic type in the *Chemical* subtree of the SN.

Thus, addition of other methods may automatically reduce most human review to deciding between Parent-Type-Missing and Child-Type-Missing. In such cases, the missing type is often simply the type of the other concept, making the correction of these unexplained **is-a** relationships relatively easy.

We found 17,022 **is-a** relationships in Meta that are unexplained by the semantic types of the concepts involved. A satisfying result would be to extend Table 1 to

⁴ Most relationships in MRREL are reciprocal, so the denominator is about half of the 10,147,419 records.

show the numbers of each of the six causes in each of the 246 relationship sets. Unfortunately, this effort is extremely difficult. Even if we had the resources to analyze each of the concept pairs manually, there are many cases where no resolution is possible without clarification from the NLM (e.g., the pair **Arteriovenous Fistula is-a Arteriovenous Malformation**). However, if the NLM were to apply our methods, they might easily resolve many of the results through editorial decisions. For example, if the NLM were to decide that, as a general editorial principle, concepts with the semantic types *Clinical Drug* or *Medical Device* should not have **is-a** relationships to concepts with semantic types in the SN's *Chemical* subtree, the causes for 14,086 (82.8%) of unexplained **is-a** relationships would be resolved.

Regardless of whether the unexplained relationships can be resolved unequivocally, our methods detect those that are inconsistent with respect to the semantic types of the concepts. Our review suggests that the majority of these inconsistent **is-a** relationships are wrong and should be deleted. We believe, therefore, that the NLM can improve the UMLS by adding our methods to the lexical [21] and semantic [9,12] auditing methods they are already using in order to identify problematic parts of Meta and SN that are deserving of human review.

6. Conclusions

The UMLS contains an enormous body of knowledge about terminologies, and its developers are expending great effort to make it coherent, consistent, and correct. Automated methods can help to focus human review on problem areas. Our method easily identifies inconsistencies in one part of the UMLS—the **parent-child is-a** relationships between concepts in Meta, as compared to the *ancestor-descendant* relationships between their corresponding semantic type, where almost one quarter are in need of correction. Our method, combined with other methods, can be applied using the UMLS developers' editorial authority to effect the necessary corrections.

Acknowledgments

This work was supported in part by Contract N01-1-3543 from the National Library of Medicine.

Appendix A

Details of the 246 unexpected sets of semantic type pairs for parent-child concept relationships in Meta.

The sets are grouped according to the parent's semantic type. For each group of sets, the code, tree ad-

dress, and name of the parent's semantic type is listed, followed by a similar listing for the child's semantic type, with one line for each set in the group. Numbers in parentheses indicate the number of concepts in each set. For example, in the first group, *Anatomical Structure* is semantic type of the parent concepts in nine sets, where the children concepts in those sets are assigned to the semantic types *Human*, *Body System*, *Body Location or Region*, *Body Space or Junction*, *Body Substance*, *Organism Attribute*, *Medical Device*, *Substance*, and *Functional Concept*, respectively. Each of the nine sets contains one parent-child concept pair.

T017: A1.2: Anatomical Structure

T016: A1.1.7.2.5.1: Human (1)

T022: A2.1.4.1: Body System (1)

T029: A2.1.5.2: Body Location or Region (1)

T030: A2.1.5.1: Body Space or Junction (1)

T031: A1.4.2: Body Substance (1)

T032: A2.3: Organism Attribute (1)

T074: A1.3.1: Medical Device (1)

T167: A1.4: Substance (1)

T169: A2.1.4: Functional Concept (1)

T018: A1.2.1: Embryonic Structure

T021: A1.2.3: Fully Formed Anatomical Structure (4)

T023: A1.2.3.1: Body Part, Organ, or Organ Component (15)

T025: A1.2.3.3: Cell (1)

T030: A2.1.5.1: Body Space or Junction (1)

T019: A1.2.2.1: Congenital Abnormality

T023: A1.2.3.1: Body Part, Organ, or Organ Component (2)

T046: B2.2.1.2: Pathologic Function (1)

T047: B2.2.1.2.1: Disease or Syndrome (3)

T190: A1.2.2: Anatomical Abnormality (4)

T191: B2.2.1.2.1.2: Neoplastic Process (1)

T020: A1.2.2.2: Acquired Abnormality

T019: A1.2.2.1: Congenital Abnormality (1)

T047: B2.2.1.2.1: Disease or Syndrome (5)

T191: B2.2.1.2.1.2: Neoplastic Process (1)

T021: A1.2.3: Fully Formed Anatomical Structure

T017: A1.2: Anatomical Structure (1)

T018: A1.2.1: Embryonic Structure (2)

T029: A2.1.5.2: Body Location or Region (7)

T031: A1.4.2: Body Substance (1)

T082: A2.1.5: Spatial Concept (1)

T104: A1.4.1.2: Chemical Viewed Structurally (1)

T123: A1.4.1.1.3: Biologically Active Substance (1)

T022: A2.1.4.1: Body System

T023: A1.2.3.1: Body Part, Organ, or Organ Component (86)

T029: A2.1.5.2: Body Location or Region (7)

- T030: A2.1.5.1: Body Space or Junction (1)
- T023: A1.2.3.1: Body Part, Organ, or Organ Component
- T017: A1.2: Anatomical Structure (1)
- T018: A1.2.1: Embryonic Structure (1)
- T019: A1.2.2.1: Congenital Abnormality (8)
- T021: A1.2.3: Fully Formed Anatomical Structure (22)
- T022: A2.1.4.1: Body System (32)
- T024: A1.2.3.2: Tissue (102)
- T025: A1.2.3.3: Cell (3)
- T026: A1.2.3.4: Cell Component (1)
- T029: A2.1.5.2: Body Location or Region (230)
- T030: A2.1.5.1: Body Space or Junction (68)
- T031: A1.4.2: Body Substance (6)
- T047: B2.2.1.2.1: Disease or Syndrome (2)
- T082: A2.1.5: Spatial Concept (6)
- T190: A1.2.2: Anatomical Abnormality (3)
- T024: A1.2.3.2: Tissue
- T017: A1.2: Anatomical Structure (1)
- T018: A1.2.1: Embryonic Structure (3)
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (174)
- T025: A1.2.3.3: Cell (2)
- T029: A2.1.5.2: Body Location or Region (1)
- T030: A2.1.5.1: Body Space or Junction (3)
- T082: A2.1.5: Spatial Concept (1)
- T025: A1.2.3.3: Cell
- T018: A1.2.1: Embryonic Structure (1)
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (7)
- T026: A1.2.3.4: Cell Component (3)
- T170: A2.4: Intellectual Product (2)
- T026: A1.2.3.4: Cell Component
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (2)
- T025: A1.2.3.3: Cell (1)
- T028: A1.2.3.5: Gene or Genome (1)
- T030: A2.1.5.1: Body Space or Junction (2)
- T075: A1.3.2: Research Device (4)
- T116: A1.4.1.2.1.7: Amino Acid, Peptide, or Protein (4)
- T123: A1.4.1.1.3: Biologically Active Substance (4)
- T029: A2.1.5.2: Body Location or Region
- T018: A1.2.1: Embryonic Structure (1)
- T021: A1.2.3: Fully Formed Anatomical Structure (3)
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (721)
- T024: A1.2.3.2: Tissue (3)
- T030: A2.1.5.1: Body Space or Junction (228)
- T033: A2.2: Finding (1)
- T082: A2.1.5: Spatial Concept (29)
- T030: A2.1.5.1: Body Space or Junction
- T018: A1.2.1: Embryonic Structure (4)
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (96)
- T024: A1.2.3.2: Tissue (2)
- T029: A2.1.5.2: Body Location or Region (261)
- T082: A2.1.5: Spatial Concept (8)
- T170: A2.4: Intellectual Product (2)
- T190: A1.2.2: Anatomical Abnormality (2)
- T031: A1.4.2: Body Substance
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (12)
- T024: A1.2.3.2: Tissue (3)
- T029: A2.1.5.2: Body Location or Region (1)
- T033: A2.2: Finding (2)
- T047: B2.2.1.2.1: Disease or Syndrome (4)
- T123: A1.4.1.1.3: Biologically Active Substance (2)
- T131: A1.4.1.1.5: Hazardous or Poisonous Substance (1)
- T184: A2.2.2: Sign or Symptom (1)
- T197: A1.4.1.2.2: Inorganic Chemical (1)
- T200: A1.3.3: Clinical Drug (4)
- T033: A2.2: Finding
- T029: A2.1.5.2: Body Location or Region (2)
- T037: B2.3: Injury or Poisoning (2)
- T040: B2.2.1.1.1: Organism Function (2)
- T046: B2.2.1.2: Pathologic Function (3)
- T047: B2.2.1.2.1: Disease or Syndrome (11)
- T169: A2.1.4: Functional Concept (5)
- T037: B2.3: Injury or Poisoning
- T033: A2.2: Finding (1)
- T046: B2.2.1.2: Pathologic Function (1)
- T047: B2.2.1.2.1: Disease or Syndrome (7)
- T040: B2.2.1.1.1: Organism Function
- T033: A2.2: Finding (1)
- T042: B2.2.1.1.2: Organ or Tissue Function (7)
- T046: B2.2.1.2: Pathologic Function (2)
- T054: B1.1.1: Social Behavior (2)
- T055: B1.1.2: Individual Behavior (1)
- T056: B1.2: Daily or Recreational Activity (3)
- T061: B1.3.1.3: Therapeutic or Preventive Procedure (1)
- T042: B2.2.1.1.2: Organ or Tissue Function
- T032: A2.3: Organism Attribute (1)
- T039: B2.2.1.1: Physiologic Function (1)
- T040: B2.2.1.1.1: Organism Function (4)
- T041: B2.2.1.1.1.1: Mental Process (2)
- T082: A2.1.5: Spatial Concept (2)
- T169: A2.1.4: Functional Concept (5)
- T046: B2.2.1.2: Pathologic Function
- T019: A1.2.2.1: Congenital Abnormality (1)
- T020: A1.2.2.2: Acquired Abnormality (3)
- T033: A2.2: Finding (1)
- T037: B2.3: Injury or Poisoning (19)
- T184: A2.2.2: Sign or Symptom (2)

- T190: A1.2.2: Anatomical Abnormality (1)
- T047: B2.2.1.2.1: Disease or Syndrome
- T019: A1.2.2.1: Congenital Abnormality (11)
- T020: A1.2.2.2: Acquired Abnormality (3)
- T031: A1.4.2: Body Substance (1)
- T032: A2.3: Organism Attribute (3)
- T033: A2.2: Finding (15)
- T034: A2.2.1: Laboratory or Test Result (1)
- T037: B2.3: Injury or Poisoning (9)
- T040: B2.2.1.1.1: Organism Function (1)
- T046: B2.2.1.2: Pathologic Function (33)
- T050: B2.2.1.2.3: Experimental Model of Disease (5)
- T067: B2: Phenomenon or Process (3)
- T080: A2.1.2: Qualitative Concept (1)
- T184: A2.2.2: Sign or Symptom (18)
- T190: A1.2.2: Anatomical Abnormality (9)
- T048: B2.2.1.2.1.1: Mental or Behavioral Dysfunction
- T033: A2.2: Finding (2)
- T047: B2.2.1.2.1: Disease or Syndrome (16)
- T067: B2: Phenomenon or Process (1)
- T184: A2.2.2: Sign or Symptom (2)
- T055: B1.1.2: Individual Behavior
- T054: B1.1.1: Social Behavior (2)
- T056: B1.2: Daily or Recreational Activity (3)
- T061: B1.3.1.3: Therapeutic or Preventive Procedure (1)
- T061: B1.3.1.3: Therapeutic or Preventive Procedure
- T062: B1.3.2: Research Activity (1)
- T074: A1.3.1: Medical Device
- T073: A1.3: Manufactured Object (7)
- T080: A2.1.2: Qualitative Concept (1)
- T104: A1.4.1.2: Chemical Viewed Structurally (1)
- T121: A1.4.1.1.1: Pharmacologic Substance (2)
- T122: A1.4.1.1.2: Biomedical or Dental Material (4)
- T167: A1.4: Substance (1)
- T077: A2: Conceptual Entity
- T017: A1.2: Anatomical Structure (1)
- T021: A1.2.3: Fully Formed Anatomical Structure (1)
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (43)
- T024: A1.2.3.2: Tissue (9)
- T040: B2.2.1.1.1: Organism Function (1)
- T042: B2.2.1.1.2: Organ or Tissue Function (1)
- T046: B2.2.1.2: Pathologic Function (1)
- T167: A1.4: Substance (1)
- T190: A1.2.2: Anatomical Abnormality (1)
- T080: A2.1.2: Qualitative Concept
- T074: A1.3.1: Medical Device (5)
- T081: A2.1.3: Quantitative Concept (1)
- T104: A1.4.1.2: Chemical Viewed Structurally (1)
- T109: A1.4.1.2.1: Organic Chemical (1)
- T121: A1.4.1.1.1: Pharmacologic Substance (1)
- T122: A1.4.1.1.2: Biomedical or Dental Material (19)
- T167: A1.4: Substance (1)
- T082: A2.1.5: Spatial Concept
- T019: A1.2.2.1: Congenital Abnormality (1)
- T023: A1.2.3.1: Body Part, Organ, or Organ Component (178)
- T024: A1.2.3.2: Tissue (2)
- T033: A2.2: Finding (1)
- T077: A2: Conceptual Entity (1)
- T104: A1.4.1.2: Chemical Viewed Structurally
- T074: A1.3.1: Medical Device (2)
- T122: A1.4.1.1.2: Biomedical or Dental Material (1)
- T125: A1.4.1.1.3.2: Hormone (1)
- T200: A1.3.3: Clinical Drug (281)
- T109: A1.4.1.2.1: Organic Chemical
- T074: A1.3.1: Medical Device (1)
- T121: A1.4.1.1.1: Pharmacologic Substance (1)
- T122: A1.4.1.1.2: Biomedical or Dental Material (6)
- T200: A1.3.3: Clinical Drug (723)
- T110: A1.4.1.2.1.9.1: Steroid
- T074: A1.3.1: Medical Device (1)
- T200: A1.3.3: Clinical Drug (197)
- T111: A1.4.1.2.1.9.2: Eicosanoid
- T200: A1.3.3: Clinical Drug (17)
- T116: A1.4.1.2.1.7: Amino Acid, Peptide, or Protein
- T074: A1.3.1: Medical Device (33)
- T087: A2.1.5.3.2: Amino Acid Sequence (1)
- T200: A1.3.3: Clinical Drug (158)
- T118: A1.4.1.2.1.8: Carbohydrate
- T200: A1.3.3: Clinical Drug (148)
- T119: A1.4.1.2.1.9: Lipid
- T122: A1.4.1.1.2: Biomedical or Dental Material (1)
- T200: A1.3.3: Clinical Drug (18)
- T120: A1.4.1.1: Chemical Viewed Functionally
- T200: A1.3.3: Clinical Drug (51)
- T121: A1.4.1.1.1: Pharmacologic Substance
- T002: A1.1.1: Plant (1)
- T031: A1.4.2: Body Substance (2)
- T073: A1.3: Manufactured Object (1)
- T074: A1.3.1: Medical Device (412)
- T109: A1.4.1.2.1: Organic Chemical (7)
- T111: A1.4.1.2.1.9.2: Eicosanoid (1)
- T119: A1.4.1.2.1.9: Lipid (5)
- T120: A1.4.1.1: Chemical Viewed Functionally (2)
- T122: A1.4.1.1.2: Biomedical or Dental Material (17)
- T123: A1.4.1.1.3: Biologically Active Substance (4)
- T129: A1.4.1.1.3.5: Immunologic Factor (1)
- T131: A1.4.1.1.5: Hazardous or Poisonous Substance (1)
- T167: A1.4: Substance (1)
- T168: A1.4.3: Food (6)

- T196: A1.4.1.2.3: Element, Ion, or Isotope (2)
 T197: A1.4.1.2.2: Inorganic Chemical (1)
 T200: A1.3.3: Clinical Drug (9296)
- T122: A1.4.1.1.2: Biomedical or Dental Material
 T074: A1.3.1: Medical Device (19)
 T104: A1.4.1.2: Chemical Viewed Structurally (2)
 T109: A1.4.1.2.1: Organic Chemical (1)
 T120: A1.4.1.1: Chemical Viewed Functionally (1)
 T121: A1.4.1.1.1: Pharmacologic Substance (4)
 T200: A1.3.3: Clinical Drug (226)
- T123: A1.4.1.1.3: Biologically Active Substance
 T087: A2.1.5.3.2: Amino Acid Sequence (1)
 T200: A1.3.3: Clinical Drug (219)
- T125: A1.4.1.1.3.2: Hormone
 T074: A1.3.1: Medical Device (31)
 T200: A1.3.3: Clinical Drug (331)
- T127: A1.4.1.1.3.4: Vitamin
 T121: A1.4.1.1.1: Pharmacologic Substance (1)
 T200: A1.3.3: Clinical Drug (582)
- T129: A1.4.1.1.3.5: Immunologic Factor
 T074: A1.3.1: Medical Device (43)
 T200: A1.3.3: Clinical Drug (131)
- T130: A1.4.1.1.4: Indicator, Reagent, or Diagnostic Aid
 T074: A1.3.1: Medical Device (122)
 T122: A1.4.1.1.2: Biomedical or Dental Material (3)
 T200: A1.3.3: Clinical Drug (286)
- T167: A1.4: Substance
 T037: B2.3: Injury or Poisoning (1)
- T168: A1.4.3: Food
 T121: A1.4.1.1.1: Pharmacologic Substance (1)
 T200: A1.3.3: Clinical Drug (17)
- T170: A2.4: Intellectual Product
 T022: A2.1.4.1: Body System (8)
 T030: A2.1.5.1: Body Space or Junction (2)
 T071: A: Entity (1)
 T074: A1.3.1: Medical Device (1)
- T184: A2.2.2: Sign or Symptom
 T033: A2.2: Finding (8)
 T046: B2.2.1.2: Pathologic Function (4)
 T047: B2.2.1.2.1: Disease or Syndrome (10)
 T048: B2.2.1.2.1.1: Mental or Behavioral Dysfunction (6)
 T067: B2: Phenomenon or Process (1)
- T185: A2.4.1: Classification
 T074: A1.3.1: Medical Device (1)
 T121: A1.4.1.1.1: Pharmacologic Substance (26)
 T122: A1.4.1.1.2: Biomedical or Dental Material (2)
 T127: A1.4.1.1.3.4: Vitamin (1)
 T130: A1.4.1.1.4: Indicator, Reagent, or Diagnostic Aid (1)
- T190: A1.2.2: Anatomical Abnormality
 T047: B2.2.1.2.1: Disease or Syndrome (2)
- T191: B2.2.1.2.1.2: Neoplastic Process
 T019: A1.2.2.1: Congenital Abnormality (1)
 T043: B2.2.1.1.3: Cell Function (1)
 T047: B2.2.1.2.1: Disease or Syndrome (7)
 T048: B2.2.1.2.1.1: Mental or Behavioral Dysfunction (2)
- T195: A1.4.1.1.1.1: Antibiotic
 T074: A1.3.1: Medical Device (2)
 T121: A1.4.1.1.1: Pharmacologic Substance (3)
 T200: A1.3.3: Clinical Drug (362)
- T196: A1.4.1.2.3: Element, Ion, or Isotope
 T200: A1.3.3: Clinical Drug (219)
- T197: A1.4.1.2.2: Inorganic Chemical
 T074: A1.3.1: Medical Device (1)
 T123: A1.4.1.1.3: Biologically Active Substance (2)
 T196: A1.4.1.2.3: Element, Ion, or Isotope (2)
 T200: A1.3.3: Clinical Drug (153)
-
- References**
- [1] Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. *JAMIA* 1998;5(1):1–11.
 - [2] National Library of Medicine. UMLS knowledge sources—experimental edition. Bethesda, MD: The Library, updated annually.
 - [3] Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. *Bull Med Libr Assoc* 1993;81(2):217–22.
 - [4] Tuttle MS, Sherertz DD, Olson NE, et al. Using META-1, the first version of the UMLS Metathesaurus. In: *Proceedings of the 14th Annual SCAMC*; 1990. p. 131–5.
 - [5] McCray AT, Hole WT. The scope and structure of the first version of the UMLS Semantic Network. In: *Proceedings of the 14th Annual SCAMC*; 1990. p. 126–30.
 - [6] McCray AT. Representing biomedical knowledge in the UMLS semantic network. In: *Proceedings of the High-Performance Medical Libraries: Advances in Information Management for the Virtual Era*; 1993. p. 45–55.
 - [7] Tuttle MS, Cole WG, Sherertz DD, Nelson SJ. Navigating to knowledge. *Methods Inf Med* 1995;34:214–31.
 - [8] Cimino JJ. Auditing the Unified Medical Language System with semantic methods. *JAMIA* 1998;5:41–51.
 - [9] Hole WT, Srinivasan S. Discovering missed synonymy in a large concept-oriented Metathesaurus. *JAMIA* 2000;7(Suppl):354–8.
 - [10] Gu H, Perl Y, Geller J, Halper M, Liu LM, Cimino JJ. Representing the UMLS as an object-oriented database: modeling issues and advantages. *JAMIA* 2000;7(1):66–80.
 - [11] Bodenreider O. Circular hierarchical relationships in the UMLS: etiology, diagnosis, treatment, complications and prevention. *Proc AMIA Symp* 2001:57–61.
 - [12] Bodenreider O. An object-oriented model for representing semantic locality in the UMLS. *Proc Medinfo* 2001;10(Pt 1):161–5.
 - [13] Cimino JJ. Battling Scylla and Charybdis: the search for redundancy and ambiguity in the 2001 UMLS Metathesaurus. *Proc AMIA Symp* 2001:120–4.

- [14] Peng Y, Halper M, Perl Y, Geller J. Auditing the UMLS for redundant classifications. Proc AMIA Symp 2002:612–6.
- [15] Gu Y, Min H, Peng Y, Zhang L, Perl Y. Using the Metaschema to audit UMLS classification errors. Proc AMIA Symp 2002:310–4.
- [16] McCray AT, Bodenreider O. A conceptual framework for the biomedical domain. In: Green R, Bean CA, Myaeng SH, editors. The semantics of relationships: an interdisciplinary perspective. Boston: Kluwer Academic Publishers; 2002. p. 181–98.
- [17] Brescia MJ, Cimino JE, Appel K, Hurwich BJ. Chronic hemodialysis using venipuncture and a surgically created arteriovenous fistula. N Engl J Med 1966;275(20):1089–92.
- [18] Halper M, Chen Z, Geller J, Perl Y. A metaschema of the UMLS based on a partition of its semantic network. Proc AMIA Symp 2001:234–8.
- [19] Zhang L, Perl Y, Halper M, Geller J, Cimino JJ. Enriching the structure of the UMLS Semantic Network. Proc AMIA Symp 2002:939–43.
- [20] McCray AT, Nelson SJ. The representation of meaning in the UMLS. Methods Inf Med 1995;34:193–201.
- [21] Bodenreider O, Burgun A, Rindfleisch TC. Assessing the consistency of a biomedical terminology through lexical knowledge. Int J Med Inf 2002;85–95.