Auditing Hierarchical Cycles to Locate Other Inconsistencies in the UMLS

Michael Halper, PhD[1], C. Paul Morrey, PhD[2], Yan Chen, PhD[3], Gai Elhanan, MD[1,5], George Hripcsak, MD[4], Yehoshua Perl, PhD[1]
[1]New Jersey Institute of Technology, Newark, NJ; [2]Utah Valley University, Orem, UT; [3]BMCC CUNY, NY, NY; [4]Columbia University, NY, NY; [5]Halfpenny Technologies, Inc., Blue Bell, PA

Abstract

A cycle in the parent relationship hierarchy of the UMLS is a configuration that effectively makes some concept(s) an ancestor of itself. Such a structural inconsistency can easily be found automatically. A previous strategy for disconnecting cycles is to break them with the deletion of one or more parent relationships—irrespective of the correctness of the deleted relationships. A methodology is introduced for auditing of cycles that seeks to discover and delete erroneous relationships only. Cycles involving three concepts are the primary consideration. Hypotheses about the high probability of locating an erroneous parent relationship in a cycle are proposed and confirmed with statistical confidence and lend credence to the auditing approach. A cycle may serve as an indicator of other non-structural inconsistencies that are otherwise difficult to detect automatically. An extensive auditing example shows how a cycle can indicate further inconsistencies.

Introduction

Concepts in the UMLS's Metathesaurus (META) [1] are created via the integration of terms from constituent source vocabularies. Synonymous terms are assigned to the same concept. Thus, a concept is an abstraction of terms with the same semantics. A UMLS relationship exists between two concepts *A* and *B* if there is a source vocabulary with such a relationship between a concept represented by a term *a* mapped to *A* and a concept represented by a term *b* mapped to *B*. A "structural inconsistency" in the META is a configuration that violates some condition that can be detected computationally. A cycle of parent relationships is an example. In [2], a process is described for eliminating cycles by deleting erroneous relationships as part of converting the UMLS into a terminological knowledge base. A cycle in the META may help identify inconsistencies among the source vocabularies and also inconsistencies arising from their integration into the META. For other problematic configurations involving inconsistencies between META hierarchical relationships and hierarchical relationships in the UMLS Semantic Network (SN) [3], see [4, 5].

In this paper, we present and evaluate a methodology for detecting the erroneous relationships in a cycle of parent relationships. An important aspect of the work is disconnecting the cycle while preserving the correct modeling and removing only relationships that represent erroneous modeling. Our emphasis is on designing computational techniques that helps to identify relationships having a high likelihood of being erroneous. Reviewing such relationships will improve the productivity of UMLS editors. Beyond exposing such an inconsistency for the sake of its own resolution, it may be an indicator of other problems involving specific concepts or their neighbors. As such, it might unearth other inconsistencies that would not otherwise be found automatically. For example, configurations involving an inconsistency between META and SN hierarchical relationships [6] can serve as indicators for locating missing hierarchical relationships in the META [7].

Bodenreider [8] explains why cycles exist in the META and offers an efficient method for detecting and removing those involving one or two concepts. Causes of cycles include differences in the granularity of the META and a source vocabulary, unspecified terms (such as "not otherwise specified" or "NOS"), metadata, compound terms, class and instance terms, implicit knowledge, and organizational conventions. Mougin and Bodenreider [9] later compared the formal method for breaking cycles of two concepts with the naïve method of cycle avoidance. The purpose was to enable work on the discovery of MeSH descriptors of a concept [8] and on the UMLS Semantic Navigator [10], a graphical interface displaying a small excerpt of the UMLS.[1] Thus, there was no particular effort in [8, 9] to disconnect erroneous relationships, and no evaluation of the correctness/incorrectness of the deleted relationships. This purely structural resolution admittedly may also be "with a significant risk of removing accurate relationships" [8]. Regarding larger cycles, Bodenreider [8] writes "The treatment of indirect circular hierarchical relationships requires a manual review of all inter-concept relationships in the cycle. No useful pattern was identified during our review."

---

[1] Bodenreider O, personal communication.

In this work, we focus on auditing cycles of three concepts, which are the simplest cases where our methodology is applicable. (It is not relevant to cycles of two concepts, as will be discussed.) The importance of the use of such a structural inconsistency for the direction of further auditing efforts is discussed. Let us note that a parent relationship is not guaranteed to represent an IS-A relationship (notated in the META with the attribute value "inverse_isa"). However, of the parent relationships with attributes, 97.4% are IS-A. Other attributes values for the parent relationship in decreasing order of percentages are: *has_part*, *has_subtype*, *has_branch*, *has_member*, *codesystem_of*, and *has_tributary*. For the parent relationships without specified attributes, the reviewer needs to take those possibilities into account when reviewing a cycle.

Extensive auditing was performed to verify the changes recommended by the application of our methodology. We note that a design principle of the UMLS is to preserve all relationships of all sources, even if their combination results in a contradictory configuration like a parent-relationship cycle. One therefore might wonder about the benefit of the suggested auditing. However, as will be demonstrated, such a configuration may be created by associating two terms having different semantics with the same concept. The association of terms with concepts is done by UMLS editors and can be modified within the framework of the UMLS mission. For example, modeling two terms of different granularity by two concepts, rather than one, will eliminate the cycle.

Methods

According to Bodenreider [8], hierarchical relationships in the MRREL file are recorded according to their origin. Relationships presented as hierarchical in the source vocabulary are recorded in the MRREL as parent/child relationships. Relationships only deemed hierarchical by the UMLS editors during the integration process are recorded as broader/narrower. In addition, some sources, like MeSH, already use *broader/narrower* relationships that are incorporated into the UMLS that way. The application of our methodology is limited in this initial study to cycles consisting only of parent/child relationships.

We identify four categories of cycles of three concepts as shown in Figure 1. Each node represents a concept and each arrow represents a parent relationship. The cycle is drawn in a clockwise direction. The category number corresponds to the number of counter-clockwise, or opposing, parent relationships present among the three concepts. We note that Category 1, 2, and 3 cycles all contain one or more cycles of two concepts, in addition to the cycle of three concepts. From now on, we will refer to a parent relationship simply as a "relationship," for brevity. A relationship that participates in a cycle of three concepts will be called a "cycle relationship" and a relationship that opposes the direction of the cycle will be called an "opposing relationship."
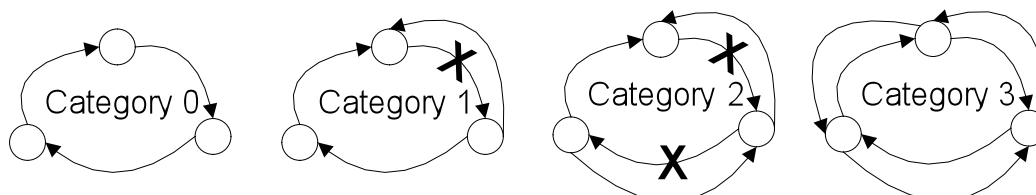

Figure 1. Categories of cycles of three concepts.

We note that this kind of analysis is not relevant for cycles of two concepts, since each of the two-cycle relationships is an opposing relationship to the other. Hence, no opposing relationships exist beyond the cycle relationship, in contrast to such phenomena for cycles of three or more concepts.

It should be noted that in a cycle of three concepts A, B, and C, the relationship (A, B) does not imply the existence or non-existence of a relationship (B, C). Similarly the relationship (A, B) does not imply the existence or non-existence of a relationship (C, A). The existence of a relationship between a pair of concepts depends on the available knowledge about the two concepts. It is independent of any connection between one of the two concepts and a third concept. Hence, there is an independence regarding the incorrectness of any two cycle relationships.

In each cycle of three concepts, at least one of the cycle relationships is incorrect. Thus, the probability of an arbitrary cycle relationship being incorrect is at least 1/3. In a Category 1 cycle, one of the cycle relationships is distinguished by having an opposing relationship. Does this cycle relationship have a higher probability of being incorrect than an arbitrary cycle relationship, since it has an explicit opposing relationship that the other cycle relationships do not? In this regard, we hypothesize:

Hypothesis **1:** For Category 1 cycles, the cycle relationship, with an opposing relationship, is most likely to be incorrect.

Again, due to the independence of the incorrectness of any two cycle relationships, the probability that at least one of two arbitrary cycle relationships is incorrect is at least 2/3. In a Category 2 cycle, two of the cycle relationships are distinguished by having opposing relationships. Is the probability that at least one such cycle relationship is incorrect higher than the probability than at least one of two arbitrary cycle relationships is incorrect? In this context, we have:

Hypothesis **2:** For Category 2 cycles, one of the two cycle relationships, with opposing relationships, is most likely incorrect.

We offer no hypothesis about Category 0 or Category 3 cycles since both are symmetric. Two random samples of Category 1 and Category 2, respectively, were provided to two auditors. For each cycle, the auditors were given a figure in which the respective preferred terms and Concept Unique Identifiers (CUIs) were displayed in concept nodes and the relationships were displayed as arrows (with their abbreviated source vocabulary identifiers listed). Definitions for the concepts were also provided when available. Figure 2 shows a sample.

Auditors also used the Neighborhood Auditing Tool (http://nat.njit.edu) [11] and the UMLS Knowledge Source Server (UMLSKS) (http://umlsks.nlm.nih.gov) to review further information from the META. Each auditor analyzed the samples independently. Disagreements were resolved by discussion until a consensus was reached. The arrived-at consensus was used as a gold standard for evaluation of the hypotheses. We used a Chi-square test to evaluate the statistical confidence for the two samples [12].

Our auditing methodology based on the ideas expressed in two hypotheses is described in the following. (The validity of the methodology will be confirmed by the validity of the two hypotheses.) It divides the treatment according to the different categories of cycles. It represents a minimal auditing effort. A thorough analysis would have the auditor consider all relationships.
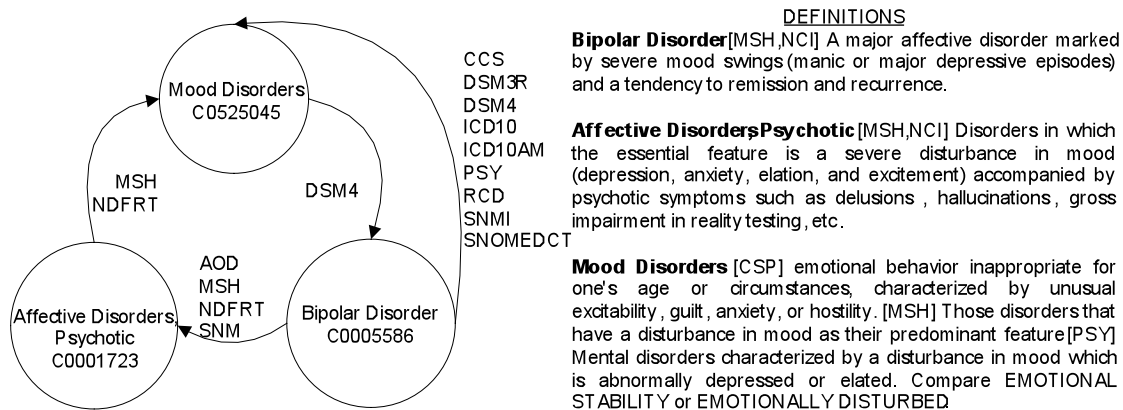

DEFINITIONS

Bipolar Disorder [MSH,NCI] A major affective disorder marked by severe mood swings (manic or major depressive episodes) and a tendency to remission and recurrence.

Affective Disorders, Psychotic [MSH,NCI] Disorders in which the essential feature is a severe disturbance in mood (depression, anxiety, elation, and excitement) accompanied by psychotic symptoms such as delusions , hallucinations , gross impairment in reality testing, etc.

Mood Disorders [CSP] emotional behavior inappropriate for one's age or circumstances, characterized by unusual excitability, guilt, anxiety, or hostility. [MSH] Those disorders that have a disturbance in mood as their predominant feature [PSY] Mental disorders characterized by a disturbance in mood which is abnormally depressed or elated. Compare EMOTIONAL STABILITY or EMOTIONALLY DISTURBED.

Figure 2. A sample figure of cycle data.

For Category 1 cycles, the auditor first considers only the cycle relationship for which an opposing relationship exists (marked with an "X" in Figure 1). Only if this relationship is deemed correct is the following cycle relationship considered. Moreover, only if the second cycle relationship is also judged correct is the final cycle relationship reviewed.

For Category 2 cycles, the auditor first considers the two cycle relationships for which opposing relationships exist (each marked with an "X" in Figure 1). Only if both cycle relationships are deemed correct is the third cycle relationship considered.

For Category 0 cycles, the auditor starts with an arbitrary relationship and continues along the cycle until one relationship is found incorrect. For Category 3 cycles, all relationships are considered.

Results

We analyzed cycles of three concepts from the UMLS version 2008AA. There are different ways to compute the cycles in the UMLS. One way is to compute the transitive closure. A cycle exists when a concept recurs. For finding all cycles for a specific length, a convenient way uses simple SQL queries. Table 1 shows the number of occurrences of cycles of three concepts divided into the four categories. One sample of 20 cycles from Category 1 and another sample of 20 cycles from Category 2 were reviewed in detail by two of the authors (YC and GE) who are trained in medicine and experienced in auditing medical vocabularies. They determined which relationship(s) and/or concept(s) should be modified for correct modeling.

Table 1. Cycles of three concepts by category.

Category	Number of Cycles	Percentage
0	148	32.45%
1	210	46.05%
2	93	20.40%
3	5	1.10%
Total:	456	100.00%

In 17 of the 20 cycles (85%) from Category 1, the auditors removed the relationship predicted to be wrong by Hypothesis 1. This is a statistically significant difference from 1/3 ($p < 0.013$). In 20 of the 20 (100%) cycles from Category 2, the auditors removed at least one of the cycle relationships predicted to be wrong by Hypothesis 2. This is a statistically significant difference from 2/3 ($p < 0.039$).

In 10 of the 20 cycles (50%) from Category 2, the auditors removed both of the cycle relationships with opposing relationships. For the sake of further testing, we conducted another study, applying our methodology to 50 more cycles of Category 1. In 42 of the cycles (84%), the consensus of the two auditors was that a cycle relationship with an opposing relationship was wrong, as predicted by Hypothesis 1. In those cases, our methodology enabled the auditor to break a cycle by reviewing only one relationship. In 6 of 50 cases (12%), the second cycle relationship was deemed to be wrong. Only in 2 cycles (4%) was the third cycle relationship judged incorrect. Using our methodology, the auditors needed to review, on average, only 1.2 relationships to break a cycle.

In this study, the auditors first used the methodology to determine how to break the cycle and then followed up with a thorough analysis of each cycle, looking for more inconsistencies. As a result of the follow-up, 34 additional cycle relationships were determined to be wrong, 11 opposing relationships were deemed wrong, and three missing relationships were noted. Furthermore, they observed other inconsistencies, such as assignment of a source term to the wrong UMLS concept, as illustrated in the extended example given below.

On average, each cycle took 40 seconds to analyze using our methodology with minimal effort, compared with an average of 188 seconds per cycle for a thorough analysis.

Extended auditing example

We present an extended audit of the cycle shown in Figure 3 (from UMLS 2008AA). Our methodology says to start with *Arteriosclerosis* (C0003850) to *Atherosclerosis* (C0004153) since it is the cycle relationship with an opposing relationship. A review of the definition of *Arteriosclerosis* shows that, in fact, "…atherosclerosis is the most common form of arteriosclerosis…," indicating that the direction of this relationship is incorrect. Further examination of the definitions shows that *Arteriolosclerosis* and *Atherosclerosis* are siblings. Therefore, the relationship from *Atherosclerosis* to *Arteriolosclerosis* ("thickening of the walls of small arteries and arterioles...") is also incorrect. In fact, this relationship was removed in 2009AA.

However, closer examination of the immediate neighborhood also reveals errors that are not directly related to the cycle itself. For example, *Arteriosclerosis* has children that are inconsistent. *Aortoiliac atherosclerosis* and *Atherosclerosis* are both children of *Arteriosclerosis* (see Figure 4(a)), while *Aortoiliac atherosclerosis* has both *Arteriosclerosis* and *Atherosclerosis* as parents. At the same time, we also find that *Arteriosclerosis* (a condition that affects only arteries) has the child *Phlebosclerosis* (a condition that affects veins and *not* arteries). Looking at the synonyms of *Arteriosclerosis*, we see "vascular sclerosis" listed. However, vascular sclerosis is a general term for sclerosis conditions of the vascular system, involving both arteries and veins. This mapping may have served as the root cause of some of the findings above. As a result, vascular sclerosis should become a separate concept from *Arteriosclerosis* and should be a parent of *Arteriosclerosis* due to its broader coverage. *Phlebosclerosis* should

become a child of the new *Vascular Sclerosis* concept and not of *Arteriosclerosis*. Figure 4(b) shows a clear hierarchical structure emerging from the changes suggested by our auditing of the initial confusing structure in Figure 4(a).
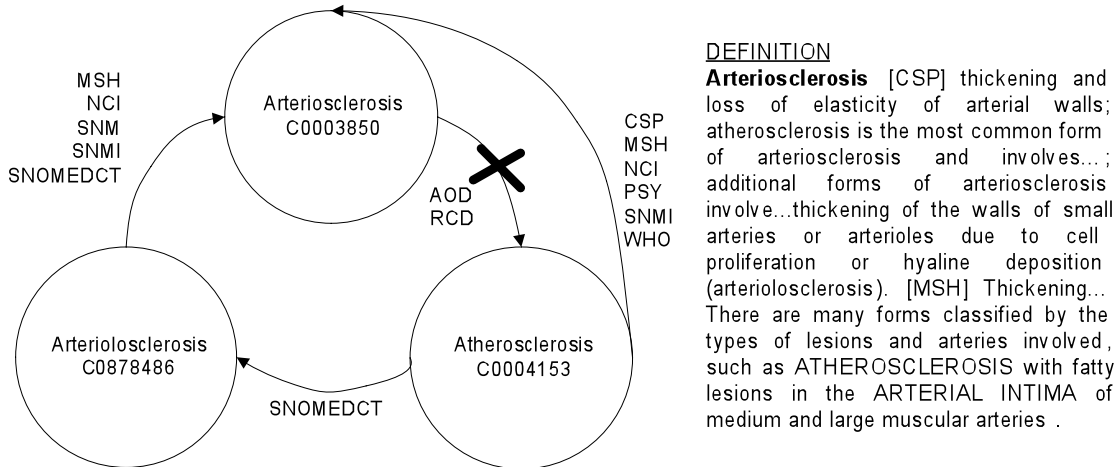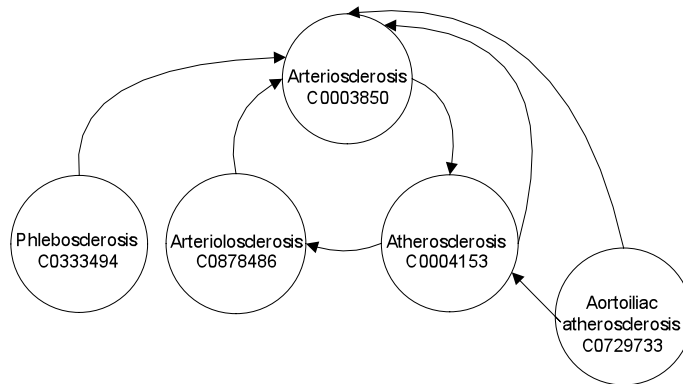

Figure 3. Cycle selected for the extended auditing example.

DEFINITION
Arteriosclerosis [CSP] thickening and loss of elasticity of arterial walls; atherosclerosis is the most common form of arteriosclerosis and involves...; additional forms of arteriosclerosis involve...thickening of the walls of small arteries or arterioles due to cell proliferation or hyaline deposition (arteriolosclerosis). [MSH] Thickening... There are many forms classified by the types of lesions and arteries involved, such as ATHEROSCLEROSIS with fatty lesions in the ARTERIAL INTIMA of medium and large muscular arteries .
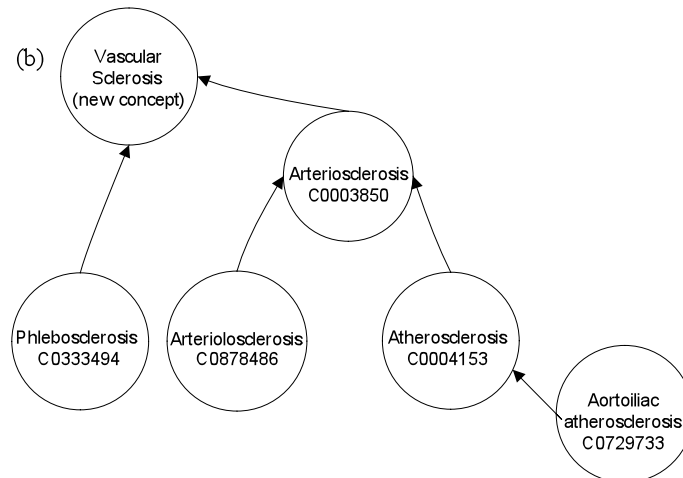

Figure 4. (a) Cycle and some neighboring concepts; (b) structure after the proposed changes.

This case demonstrates the general phenomenon of a cycle indicating some modeling confusion or contradicting perceptions, which will be manifested as errors in the neighborhood of the cycle. The auditor will use a cycle to trigger auditing of the neighborhood elements to find errors that are difficult—or perhaps impossible—to detect automatically.

The cycle of Figure 3 no longer exists in the UMLS. Changes in the SNOMED CT source vocabulary were propagated into the UMLS, causing the removal of one of the incorrect cycle relationships, from *Atherosclerosis* to *Arteriolosclerosis*. We note that although the cycle of three concepts was broken, the issues we present in the extended example regarding the marked relationship from *Arteriosclerosis* to *Atherosclerosis* and the missing concept *Vascular Sclerosis* still exist. We note also that this marked relationship, with sources AOD and RCD, was already deleted from SNOMED when RCD was integrated into it, showing the agreement of SNOMED editors with our recommendation. Another of our recommendations, the removal of the parent relationship from *Aortiliac atherosclerosis* to *Arteriosclerosis*, in the context of the UMLS 2008AB version was likewise propagated into 2010AB due to changes in SNOMED.

Discussion

There are various ways of practicing auditing. One extreme is the ideal scenario where the auditor can pay attention to all relevant details available and explore the possibility of a discovered inconsistent configuration as an indicator to more inconsistencies or errors. In the other extreme, the auditor has a quota of concepts to review in a given time and needs to optimize his productivity. In that case, the auditor tries to view only necessary knowledge to determine how to resolve an inconsistent configuration, but has no time to explore other potential problems in the neighborhood. In this paper, we tried to address the needs of both options. For the ideal scenario, we showed how a cycle of three concepts may indicate more modeling problems, which otherwise would have been difficult to expose. That is, automatically detectable, inconsistent configurations help to expose more problems that are not necessarily automatically detectable. For an auditor who can afford only a minimal effort in resolving a cycle, we offer a methodology where, in most cases, only one or two cycle relationships need to be reviewed. In such a setting, it would be beneficial to keep records of the changes in the hope of later exploration of hidden inconsistencies. In fact, the cycle we used in the extended example demonstrates the potential value of applying this method to previous UMLS revisions to identify problems that would otherwise be missed.

Hierarchical cycles of two or more concepts represent structural ambiguity. Controlled medical vocabularies avoid such ambiguity by adopting a directed acyclic graph (DAG) infrastructure. The UMLS, however, is not a controlled terminology but a terminological system that is fed from its many source vocabularies, adapting the content of META to their many structures. Consequently, cycles can be formed as a side-effect of the integration process itself.

While cycles typically do not exist in an individual UMLS source, the occurrence of a cycle in the UMLS offers an opportunity to detect potential source errors that would not otherwise be automatically detectable within the context of the source itself. Hence, in addition to the benefits to the UMLS, we gain the added benefit of corrections in source vocabularies observed only when considering contradictions arising as a result of integration with other sources in the scope of the META.

The typical non-reflexive META hierarchical cycle that involves more than one source vocabulary may present significant challenges to external applications. Cycles are inherently problematic for applications designed to traverse the ancestor-descendant pathways within the META and build upon the semantics represented within it. Consider the three concepts *A*, *B*, and *C*, where *A* is parent of *B*, *B* is parent of *C*, and also *C* is parent of *A*. If *C* is given as a starting point for an application designed to collect more granular concepts, *A* may be erroneously included. A simple such example can be the case of an application that attempts to display dynamic content for drop-down menus for further refinement based on META's hierarchy, such as in many semi-structured or structured data entry screens. If concept *A* were included in one of the drop-down selections, it may present a confusing scenario for the user. If selected, it could result in outright erroneous data entry.

For example, the UMLS concept *Bipolar disorder* is both a parent and a child of *Mood disorders*, as well as a child of *Affective disorders*, *psychotic* (see Figure 2 above). In the abovementioned scenario, an application that wishes to allow users to select more specific types of bipolar disorders may actually present its users with *Involutional depression*, which is not a type of bipolar disorder. This incorrect presentation results from the fact that *Involutional depression* is a child of *Major depressive disorder*, which in turn is a child of *Mood disorders*. But due to the latter being a child of *Bipolar disorder* via the cycle, *Involutional depression* is also *Bipolar disorder*'s descendant. Our theory implies that *Bipolar disorder* being a parent of *Mood disorders* is likely to be incorrect. Indeed, removing this

relationship from the UMLS would eliminate *Involutional depression* from the returned specific types of bipolar disorders.

Limitations and Future Work

This paper describes an initial study, with the purpose of examining whether the existence of opposing relationships for the cycle can function as an indicator of which cycle relationships are likely to be erroneous. Cycles of three concepts are the smallest cycles for which opposing relationships can exist, since, as noted, for a cycle of two concepts, the opposing relationships are not distinguishable from the cycle relationships themselves.

A Category-0 cycle has no opposing relationships, so we examined other computational solutions to identify relationships likely to be erroneous. We considered the idea of voting by the number of UMLS source vocabularies for each cycle and opposing relationship. This was used by Bodenreider in a procedure for dealing with two-concept cycles [8]. The assumption is that the relationships voted on by more sources are correct, and the ones voted on by less are erroneous. It could be used, for example, to indicate the erroneous cycle relationship in Figure 3. Another tack would be to give more weight to sources known to be relatively reliable ("weighted voting"). We experimented with both approaches for a sample of 20 cycles of Category 0. Neither approach was successful when compared to the auditing done by our domain experts. The voting approach may work for cycles of two concepts, but it was not actually evaluated for correctness [8]. Such an evaluation is planned for future work.

The success of our reported approach suggests that it be extended to larger cycles (of four or more concepts). Note that for larger cycles, the number of categories not only grows linearly, but they become more complex. For example, for Category-2 cycles of four relationships, there are two configurations, one with two consecutive opposing relationships and one with "alternating" opposing relationships. Future research is planned to study the adaptation of our approach to larger cycles.

Let us note that the number of cycles decreases as the length (i.e., number of concepts in the cycle) increases. For example, in the 2011AA release of the UMLS, there are 1,668 cycles of length two, 433 cycles of length three, 196 cycles of length four, and 89 cycles of length five. Hence, developing a technique for indicating the relationships of a cycle likely to be in error, and thus saving on auditing efforts, is really only needed for a few additional lengths, say, four, five, and perhaps six. Beyond that length, the number of cycles decreases to an amount that will just require individual review, not an automated approach.

Furthermore, one needs to consider cycles consisting only of "broader" relationships and mixed cycles consisting of both parent and broader relationships. It should be investigated whether similar hypotheses hold up for these other possibilities of length and relationship type. Statistics need to be collected regarding the percentages of cycles that are indicators of other problems and what kinds of problems are frequently observed.

Of the 90 cycles we analyzed using the UMLS 2008AA, 81 are still present in the UMLS 2010AB. We observe that the changes in the other nine cycles are due to modifications in the source vocabularies that propagated into the META. Only in two of the nine cases does the modification agree with our recommendation. Hence, for the other cases, the wrong relationship was not detected and in the newer revisions the problem will not be detectable automatically.

We recently reported these findings to the NLM. By UMLS policy, the content of a source vocabulary is preserved—even in cases of errors and contradictions. In the case of a cycle caused by a contradiction between sources, a change can only be achieved in the UMLS by communicating the problem to the respective authoritative organization in charge of the "offending" source. Only later will the change be propagated into the UMLS. But in some cases, the cycles and incorrect relationships were caused by choices of mapping or lower UMLS granularity compared to the source. In such cases, the UMLS team can correct the modeling to disconnect the cycles and avoid the erroneous relationships. We see such an example in Figure 2, where inverse relationships between *Mood Disorders* and *Bipolar Disorder* are both coming from the DSM4 source vocabulary. However, examining the modeling of DSM4 atoms in the UMLS reveals that *Bipolar Disorder* is a child of *Mood Disorders* and a parent of *Mood Disorder NOS*. The cycle was caused by the UMLS modeling *Mood Disorder NOS* as synonym of *Mood Disorders*. As a matter of fact, the notion of "NOS concept" has been given as one of the reasons for cycles in the UMLS [8].

Conclusion

Cycles of parent relationships represent contradictory modeling within the META of the UMLS. Their causes are manifold and their resolution is important for a variety of reasons. We have introduced a methodology for auditing

of cycles that seeks to discover and delete erroneous relationships only. It was applied to cycles comprising three concepts. Hypotheses were tested to confirm the effectiveness of our approach. Overall, the methodology helps minimize the auditing resources required to resolve cycles. It was also shown that cycles serve as good indicators of the presence of other inconsistencies in the META.

References

1. UMLS – Metathesaurus. http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html. Accessed: March 16, 2011.
2. Schulz S, Hahn U. Medical knowledge reengineering – converting major portions of the UMLS into a terminological knowledge base. Int'l Journal of Medical Informatics. 64(2-3): p. 207–21, Dec. 2001.
3. The UMLS Semantic Network. http://semanticnetwork.nlm.nih.gov. Accessed: March 16, 2011.
4. Cimino JJ, Min H, Perl Y. Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus. JBI. 36(6): p. 450–61, Dec. 2003.
5. Geller J, Morrey CP, Xu J, Halper M, Elhanan G, Perl Y, Hripcsak G. Comparing inconsistent relationship configurations indicating UMLS errors. Proc. 2009 AMIA Annual Symposium, p. 193–7.
6. Chen Y, Gu H, Perl Y, Geller J, and Halper M. Structural group auditing of a UMLS semantic type's extent. JBI, 42(1): p. 41–52, Feb. 2009.
7. Chen Y, Gu HH, Perl Y, and Geller J. Structural group-based auditing of missing hierarchical relationships in UMLS. JBI, 42(3): p. 452–467, June 2009.
8. Bodenreider O. Circular hierarchical relationships in the UMLS: Etiology, diagnosis, treatment, complications and prevention. Proc. 2001 AMIA Annual Symposium.
9. Mougin F, Bodenreider O. Approaches to eliminating cycles in the UMLS Metathesaurus: naïve vs. formal. Proc. 2005 AMIA Annual Symposium.
10. Bodenreider O. A semantic navigation tool for the UMLS. Proc. AMIA 2000 Annual Symposium, p. 971.
11. Morrey CP, Geller J, Halper M, Perl Y. The Neighborhood Auditing Tool: a hybrid interface for auditing the UMLS. JBI. 42(3): p. 468–89, June 2009.
12. Snedecor GW, Cochran WG. Statistical Methods. 8th ed. Ames, Iowa: Iowa State University Press; 1989.