

Review

A review of auditing techniques for the Unified Medical Language System

Ling Zheng^{1,*}, Zhe He^{2,*}, Duo Wei³, Vipina Keloth⁴, Jung-Wei Fan⁵, Luke Lindemann⁶, Xinxin Zhu⁶, James J. Cimino ⁷, and Yehoshua Perl⁴

¹Department of Computer Science and Software Engineering, Monmouth University, West Long Branch, New Jersey, USA, ²School of Information, Florida State University, Tallahassee, Florida, USA, ³School of Business, Stockton University, Galloway, New Jersey, USA, ⁴Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey, USA, ⁵Division of Digital Health Sciences, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA, ⁶Center for Biomedical Data Science, Yale School of Medicine, New Haven, Connecticut, USA, and ⁷Informatics Institute, University of Alabama at Birmingham, Birmingham, Alabama, USA

*These authors contributed equally.

Corresponding Author: Zhe He, PhD, School of Information, College of Communication and Information, Florida State University, 142 Collegiate Loop, Tallahassee, FL 32306-2100, USA; zhe@fsu.edu

Received 10 February 2020; Revised 5 May 2020; Editorial Decision 11 May 2020; Accepted 13 May 2020

ABSTRACT

Objective: The study sought to describe the literature related to the development of methods for auditing the Unified Medical Language System (UMLS), with particular attention to identifying errors and inconsistencies of attributes of the concepts in the UMLS Metathesaurus.

Materials and Methods: We applied the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach by searching the MEDLINE database and Google Scholar for studies referencing the UMLS and any of several terms related to auditing, error detection, and quality assurance. A qualitative analysis and summarization of articles that met inclusion criteria were performed.

Results: Eighty-three studies were reviewed in detail. We first categorized techniques based on various aspects including concepts, concept names, and synonymy ($n = 37$), semantic type assignments ($n = 36$), hierarchical relationships ($n = 24$), lateral relationships ($n = 12$), ontology enrichment ($n = 8$), and ontology alignment ($n = 18$). We also categorized the methods according to their level of automation (ie, automated systematic, automated heuristic, or manual) and the type of knowledge used (ie, intrinsic or extrinsic knowledge).

Conclusions: This study is a comprehensive review of the published methods for auditing the various conceptual aspects of the UMLS. Categorizing the auditing techniques according to the various aspects will enable the curators of the UMLS as well as researchers comprehensive easy access to this wealth of knowledge (eg, for auditing lateral relationships in the UMLS). We also reviewed ontology enrichment and alignment techniques due to their critical use of and impact on the UMLS.

Key words: quality assurance, unified medical language system, auditing, review

INTRODUCTION

The Unified Medical Language System (UMLS)^{1–6} is a unique system designed by the National Library of Medicine (NLM), spear-

headed by Donald Lindberg, Betsy Humphreys, and Alexa McCray, to integrate a large collection of biomedical terminology and ontology sources (currently 213 [<https://www.nlm.nih.gov/research/umls/>]

Table 1. Selection criteria for article inclusions

Type	Criteria	Rationale
Inclusion criteria	Methods for finding errors or inconsistencies of aspects of UMLS concepts	Errors and inconsistencies of concept names, synonyms, ST assignments, hierarchical (IS-A) relationships, and lateral relationships.
	UMLS auditing tools, surveys, and auditors' performance	Owing to their relevance for the auditing process.
	Auditing observed during the integration of sources into the UMLS	Limiting the review to this side effect of the integration.
	Topological patterns techniques and alignment techniques for enhancement of the UMLS sources	Owing to their major use of the UMLS although their purpose is to enhance UMLS sources; The enhancement will indirectly be leading to modifications of the UMLS. In addition, identifying missing synonyms for UMLS concepts is another byproduct of these techniques.
Exclusion criteria	Coverage of the UMLS	Assessing the coverage of the UMLS concepts is not relevant to QA.
	Applications of the UMLS	Applications of the UMLS such as information retrieval or natural language processing are not relevant to QA.
	Auditing of sources of the UMLS	Auditing the sources is not relevant to UMLS QA.
	Integration of sources into the UMLS	Integration of the sources into the UMLS is not relevant if no auditing of the UMLS is observed.
	Refinements, extensions, or summarization networks of the UMLS SN	Refinement, extension, partition, and summarization of the UMLS SN are not focused on QA of UMLS concepts.
	Not related to UMLS (eg, UML)	Some articles that are irrelevant to the UMLS were retrieved by PubMed search (eg, Unified Modeling Language).
General UMLS article not relevant to QA	Some general UMLS development articles were retrieved by PubMed search.	
Not an article	Conference abstracts are excluded.	

QA: quality assurance; SN: Semantic Network; ST: semantic type; UML: Unified Modeling Language; UMLS: Unified Medical Language System

sourcereleasedocs/index.html]) into a Metathesaurus. In the UMLS, synonymous terms from multiple sources are mapped to the same UMLS concept; each concept is classified as belonging to 1 or more of 127 semantic types (STs), taken from the UMLS Semantic Network (SN).⁷⁻⁹ The SN also includes 54 semantic relations (SRs) that indicate potential relationships among concepts based on their STs (counts taken from file SRDEF, archived at https://semanticnetwork.nlm.nih.gov/download/sn_current.tgz).

Integrating 213 biomedical sources of various models and naming standards poses difficulties; errors and inconsistencies are inevitable.¹⁰ Starting with Cimino,¹¹ many researchers have designed techniques for “auditing the UMLS” (ie, finding and categorizing errors and inconsistencies of the UMLS). The purpose of this article is to review and categorize these various auditing techniques, also known as quality assurance (QA) techniques, to summarize the wealth of experience applied to this task.

A general UMLS users study by Chen et al¹² reported concerns about errors in UMLS concepts, particularly with hierarchical relationships. A special issue of the *Journal of Biomedical Informatics* on methods for auditing biomedical terminologies¹³ contained a review article¹⁴ in which 51 of the cited articles described work involving the UMLS. This was more than for any single terminology, demonstrating the importance researchers attribute to the UMLS. This importance stems from the unique design of the UMLS, which enabled the rich body of research reviewed in this survey. Furthermore, it opened the possibility of comparing and contrasting multiple UMLS source terminologies. In that article, Zhu et al¹⁴ described articles according to several dimensions, including quality factors, knowledge source, automation level, and aspects of terminology content. Amith et al¹⁵ presented a review article of general ontology evaluation techniques, which is not a systematic review. In addition, some approaches that are appropriate for evaluating single ontologies, for example, “compare the target ontology to a ‘gold stand-

ard,”¹⁶ as outlined in Amith et al,¹⁵ are at best only partially applicable to the UMLS due to its unique content (ie, integration of many source vocabularies) and unique purpose of serving as multi-purpose middleware for a wide range of different applications and systems. It has 10 UMLS references but only few are discussed.¹⁷⁻²⁶

In contrast to Zhu et al,¹⁴ this article considers auditing of just the UMLS and is restricted to methods for identifying errors and inconsistencies in the various aspects of the Metathesaurus concepts. We include alignment and topological pattern enhancements of the UMLS sources due to the intensive use of the UMLS as a matching intermediary. Furthermore, enhancements to the UMLS sources have an indirect impact on future releases of the UMLS. To limit the scope of this review, we did not consider refinements, extensions, partitions and summarization of the SN, which were reviewed by Zhu et al.¹⁴ Out of the 51 UMLS references in Zhu et al,¹⁴ only 23^{11,27-48} conformed to the strict interpretation of “auditing the UMLS” used in this review. Table 1 provides the criteria used for inclusion and exclusion of articles considered in this review.

Our study concentrates on the methodology of the auditing techniques. Furthermore, our review categorizes studies differently from Zhu et al,¹⁴ in which the major categorization is by quality factors and levels of automation. In this review, the 83 studies are categorized according to audited aspects of the concepts, which provides a clear and comprehensive picture of methodologies for UMLS auditing.

We classify techniques based on particular concept characteristics: names, synonyms, ST assignments, hierarchical (IS-A) relationships, and lateral relationships. Figure 1 shows the UMLS interface for the concept *Bipolar Disorder* illustrating the various aspects. For each article, we provide a brief description of its technique(s), identifying the audited aspect(s), the degree of manual versus automated approach, and the source of knowledge used to support the technique. Results appear in the [Supplementary Appendix](#).

The screenshot displays the UMLS Terminology Services Metathesaurus Browser. At the top, it identifies itself as a service of the U.S. National Library of Medicine | National Institutes of Health. The main header reads 'UMLS Terminology Services Metathesaurus Browser'. Below this is a navigation bar with links: UTS Home, About, Browse, Download, APIs, Tools, Help.

The interface is divided into several sections:

- Search Section:** Includes a search bar with the text 'Try the new Metathesaurus Browser (Alpha Release)'. Below it are radio buttons for 'Term', 'CUI', and 'Code'. A search box contains 'Bipolar Disorder' and a 'Go' button. There are dropdown menus for 'Release' (set to '2019AB'), 'Search Type' (set to 'Word'), and 'Source' (set to 'All Sources').
- Search Results (321):** A list of search results for 'Bipolar Disorder', showing various CUIs and their corresponding terms, such as 'C0005586 Bipolar Disorder', 'C0236788 Bipolar II disorder', and 'C0270419 Mild bipolar disorder'.
- Basic View Section:** Contains detailed information for the selected concept:
 - Concept:** [C0005586] Bipolar Disorder
 - Semantic Type:** Mental or Behavioral Dysfunction [T048]
 - Definition:** (Empty)
 - Synonyms (83):** A list of 83 synonyms from different sources, including 'Affective Psychosis, Bipolar', 'BIPOLAR AFFECTIVE DISORDER', 'BIPOLAR DISORDER', 'BIPOLAR DISORDERS', 'BPAD', 'Bi-polar Disorder', 'Bipolar Affective Disorder', 'Bipolar Affective Psychosis', 'Bipolar Disorder', 'Bipolar Disorders', 'Bipolar Mood Disorder', 'Bipolar affective disorder', 'Bipolar affective disorder, unspecified', 'Bipolar disorder', 'Bipolar disorder (disorder)', 'Bipolar disorder NOS', 'Bipolar disorder, NOS', 'Bipolar disorder, unspecified', 'Bipolar disorders', 'Depression;manic', 'Disorder, Bipolar', 'Disorder;bipolar', 'MANIC DEPRESSIVE ILLNESS', 'MANIC DEPRESSIVE REACTION', 'MDI - Manic-depressive illness', 'Manic Depression', 'Manic Depressive Disorder', 'Manic Depressive Psychosis', 'Manic depression', 'Manic depressive', 'Manic depressive illness', 'Manic depressive psychosis', 'Manic depressive reaction', 'Manic-Depression', and 'Manic-Depressive Illness'.
 - Relations (1691):** A list of 1691 relationships between the concept and other concepts. The first few are: 'PAR | | SNMI | Mood Disorders | C0525045', 'PAR | inverse_isa | SNOMEDCT_US | Mood Disorders | C0525045', 'PAR | inverse_isa | SCTSPA | Mood Disorders | C0525045', 'PAR | | ICD10AM | Mood Disorders | C0525045', 'PAR | inverse_isa | NCI | Mood Disorders | C0525045', 'PAR | | CCS | Mood Disorders | C0525045', 'PAR | inverse_isa | HPO | Manic mood | C0564408', 'PAR | | MDRSPA | Manic and bipolar mood disorders and disturbances | C0851581', 'PAR | | MDRCZE | Manic and bipolar mood disorders and disturbances | C0851581', and 'PAR | | MDRJPN | Manic and bipolar mood disorders and disturbances | C0851581'.

Figure 1. The Unified Medical Language System (UMLS) Metathesaurus Browser user interface, displaying information for the concept *Bipolar Disorder*. The interface shows the focus concept *Bipolar Disorder* at the top of the right box, followed by the semantic type of the concept and 83 synonyms from different sources (out of which only 36 fit on the screen). Relationships (including hierarchical, lateral, and qualifiers) between *Bipolar Disorders* and 1691 (not necessarily different) target concepts are listed below the synonyms (shown to the right of synonyms in this figure), showing the relation, relationship attribute, source terminology, the term name in the source terminology, and the concept unique identifier (CUI) for each related concept. For example, *Mood Disorders* appears 6 times, each mapped to the same CUI, because this relationship is found in 6 source terminologies. The screenshot was taken on January 31, 2020, using UMLS version 2019AB.

MATERIALS AND METHODS

Identifying the references

To identify relevant articles, we followed the Institute of Medicine's standards for systematic review and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).⁴⁹ Our process consisted of 4 steps: (1) identifying relevant keywords; (2) formulating the search query to identify relevant articles from PubMed and Google Scholar; (3) screening titles based on predefined inclusion and exclusion criteria; and (4) reviewing abstracts and full texts to exclude irrelevant articles and code for reasons. For details on the processing of the steps see [Figure 2](#) and the [Supplementary Appendix](#).

Errors in the UMLS

Errors may occur with every aspect of a concept in the Metathesaurus. We distinguish between 2 kinds of errors. The first is errors that

are imported to the UMLS from source terminologies. Sometimes these errors are invisible within the source terminology and are exposed in the UMLS as part of an illegal structure (eg, a cycle of IS-A relationships).

The second kind of errors are "made in the UMLS." One example is the ST assignment. This is a unique feature of the UMLS, in which editors assign STs to concepts. Another example is the identification of concepts from different terminologies. In the process of integrating a source into the UMLS (eg, Systematized Nomenclature of Medicine [SNOMED]⁵⁰ or Gene Ontology [GO])⁵¹ editors must establish new concepts that do not yet appear in the Metathesaurus. When this concept appears later in another terminology with the same name, it is associated with the same concept. If this concept appears in another terminology with a different name, it is assigned as a synonym for this concept. However, those decisions are not always simple. Different terminologies have various naming conventions. There are also cases of homonyms, in which different

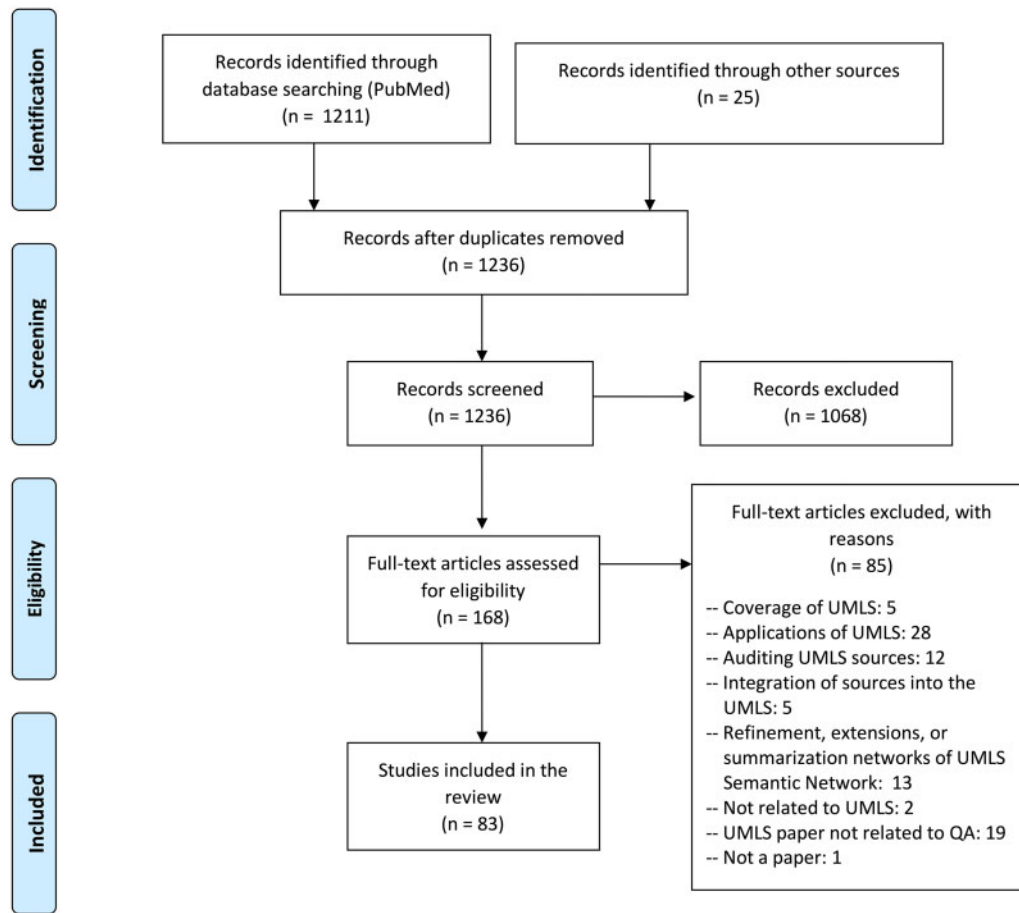


Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart for identifying the articles to be included in this review.

terminologies use the same name with varying semantics. Terminologies sometimes have conflicting views about which of 2 concepts is more specific. Thus, UMLS editors sometimes erroneously multiply singular concepts or unify multiple concepts. Such errors can also cause cycles in the UMLS.

Auditors of the UMLS should identify whether the error is made in the UMLS and should be reported to the NLM, or whether the error is in one source terminology and should be reported to its curators. When the error is corrected in the source terminology, it will disappear from the next release of the UMLS.

An interesting phenomenon is that sometimes several errors come together for 1 or several similar concepts. For example, an auditor looking for an explanation for a missing lateral relationship may discover a wrong or missing IS-A relationship. Hence, detecting 1 error may propagate the detection of more errors that would otherwise be hidden.

Categorization of the articles

We first categorized and discussed the 83 included articles based on the aspects that they focus on (1) concepts, concept names, and synonyms; (2) semantic type assignments; (3) hierarchical (IS-A) relationships; (4) lateral relationships; (5) topological pattern-based ontology enrichment; and (6) ontology alignment. Then we coded the automation level (ie, automated systematic, automated heuristic, or manual) and knowledge source of the techniques (ie, intrinsic knowledge, extrinsic knowledge, or combined intrinsic and extrinsic

knowledge), based on the definitions of the characteristics (shown in Table 2) following Zhu et al.¹⁴

Categorizing the articles according to these aspects allows us to combine the description of similar studies and present in our review the progress of the research ideas underlying them (see Discussion). The descriptions are succinct and focus on the essence of the techniques used, rather than the results, in order to cover a large number of studies in a limited space. The auditing results appear in the Supplementary Table 1.

RESULTS

Auditing techniques for the various aspect categories

Concepts, concept names, and synonyms

Synonym detection serves as a critical check on preventing the creation of redundant concepts when the UMLS intakes updates from source terminologies. Redundant (synonymous) concepts were detected lexically by Cimino,^{11,27} by looking for pairs of concept names with the same words in a different order or with different punctuation. Those comparisons were expanded by identifying interchangeable keyword synonyms. Hole and Srinivasan²⁸ introduced several heuristics used by the NLM to boost the sensitivity: lexical tweaks (eg, trimming space or punctuation), synonymous word swaps (eg, “renal” vs “kidney”), and enhanced matching through synonymous token discovery (originally credited to R.A. Miller). Huang et al²⁹ further expanded the third heuristic into a

Table 2. Definitions of the characteristics of the auditing techniques

Characteristics of the auditing technique	Types	Definitions
Automation level	Automated systematic	Automated systematic methods are implemented as rule-checking programs or algorithms that can automatically identify potential errors and inconsistencies in the terminology.
	Automated heuristic	Automated heuristic methods are based on rules that make inferences about terminology content and seek to identify those inferences to find likely errors and inconsistencies in the terminology.
	Manual	Manual review relies on a terminology reviewer (often a domain expert) to manually audit a certain aspect(s) of a terminology, with or without the support of a computerized user interface.
Knowledge source	Intrinsic knowledge	Intrinsic knowledge is the information derived from the classification scheme, hierarchy, relationships, or other attributes present within the terminology itself.
	Extrinsic knowledge	Extrinsic knowledge is derived from an outside source, such as other terminologies or human expert knowledge.

formal algorithm (GSMake). The assumption is that nonoverlapping word(s) from 2 overlapping synonyms of the same concept unique identifier (CUI) may be collected as interchangeable components to facilitate sensitive matching. As an extension, Huang et al³⁰ substituted the GSMake for WordNet *synsets* in generating the alternative match terms. They came up with ancillary heuristics to control the exploding variants by the maximum number of allowed word swaps per term and the maximum length of a term being processed. Huang et al⁵² applied these culminated methods to audit duplicate concepts that had been incorporated from the SNOMED Clinical Terms (CT) into the UMLS.

Bodenreider and McCray⁵³ and McCray et al⁵⁴ aggregated the STs into 15 semantic groups. Erdogan et al^{55,56} discovered missing concepts by identifying concepts that hierarchically belonged to 2 inconsistent semantic groups and yet had no ancestors carrying such inconsistency. They implemented an answer set programming method to enhance efficiency. They found that an additional concept sharing synonym at the inconsistent spot must be created to untangle the 2 semantic lineages, a common type of error.

The coverage of UMLS genomic and proteomic concepts and relationships is compared with terms from GO and LocusLink.³¹ The relationships are well-represented but the coverage of fine-grained concepts is limited (year of study: 2002). Disambiguation of ambiguous terms is recommended with a systematic polysemy focus using context rather than integrating gene and gene product names.

Liu et al⁵⁷ performed a systematic analysis of abbreviations in the UMLS synonyms. The prevalent ambiguity (1 abbreviation shared by multiple CUIs) was not necessarily a quality indicator, but the evaluation of abbreviation used in clinical text could be viewed as a benchmark of lexical completeness for representing the medical domain.

Merrill⁵⁸ applied a formal semantic analysis to clarify fundamental notions underlying the UMLS: atom, term, and concept. He proposed an approach based on “synonymy-based Metathesaurus models,” with theoretical principles for maintaining the UMLS semantic integrity.

Semantic type assignment

The UMLS consists of 2 levels: the Metathesaurus and the SN of 127 STs.^{59,60} Each Metathesaurus concept is assigned 1 or more STs. In Geller et al³⁹ and Gu et al,^{40,41} as an outgrowth of the development of an object-oriented database version of the UMLS, Gu et al^{40,41} introduced the Refined SN (RSN) Abstraction Network,⁶¹ in which each concept is assigned only 1 Refined ST (RST). The intersection of multiple STs is described by an Intersection ST (IST).

Concepts assigned ISTs are more complex, with multiple semantics and a higher probability of errors. For example, Cimino¹¹ found concepts assigned to mutually exclusive STs to be contradictions. Figure 3 illustrates an IST and the mapping between the 2 levels of the UMLS.

Consider an IST $A \cap B$ such that A is an ancestor of B in SN. According to the specificity rule,⁶⁰ the assignment of A is redundant and the concepts should be assigned only B. Peng et al⁴⁶ designed an algorithm to find and remove all redundant ST assignments, which was implemented by Srinivasan of UMLS and is used in the UMLS production (YP personal communication with Suresh Srinivasan in 2005).

Cimino^{11,27} used multiple ST assignments to identify potentially ambiguous UMLS “concepts” with multiple meanings. ST assignments were also used to identify inconsistent hierarchical relationships where a type assigned to a child concept was neither identical to, nor a descendant of any of the types assigned to its parent. In 2 versions of the UMLS, inconsistent classifications were detected. Cimino et al⁴² refined the use of ST assignments to detect and classify errors in hierarchical relationships, finding concept pairs with inconsistent classification.

Gu et al^{43,44,62} showed that uncommonly modeled small IST concepts had higher error rates than larger ones did. ST assignment errors were found in Gu et al.⁴⁴ For example, *Scotch Tape Mount*, a laboratory procedure for detecting pinworms, was assigned to Bacterium and Laboratory Procedure. A broader study of all 232 such concepts⁶² by 4 domain experts showed that multiple auditors are required to achieve reliability in auditing complex IST concepts. They experimented auditing with the Neighborhood Auditing tool (NAT) (Figure 4).⁶³ A study found it more effective than the UMLS browser (Figure 1). Ochs et al⁶⁵ extended the NAT for a relationship-centric browsing and auditing tool.

The cohesive meta-schema^{66,67} is a partition of the STs of SN into the Meta Semantic Types (MST). In auditing concepts of small pure MSTs,⁴⁵ a higher error rate is found because concepts in the intersection of different MSTs are more likely to have errors. Mouglin et al¹⁸ analyzed concepts with multiple semantic groups in the UMLS. Categorization inconsistency between parent and child concepts is an indicator of categorization error.

Chen et al^{37,68} considered group auditing by expanding RST extent (the set of concepts assigned a RST). For each RST, an envelope of parents and children of extent concepts is defined. The expansion algorithm iteratively suggests concepts for review by a domain expert. Analyzing the results, Chen et al⁶⁹ reported on additional concepts missing ST that were not found because some concepts are assigned ST *Classification*, which blocks the expansion. A revised al-

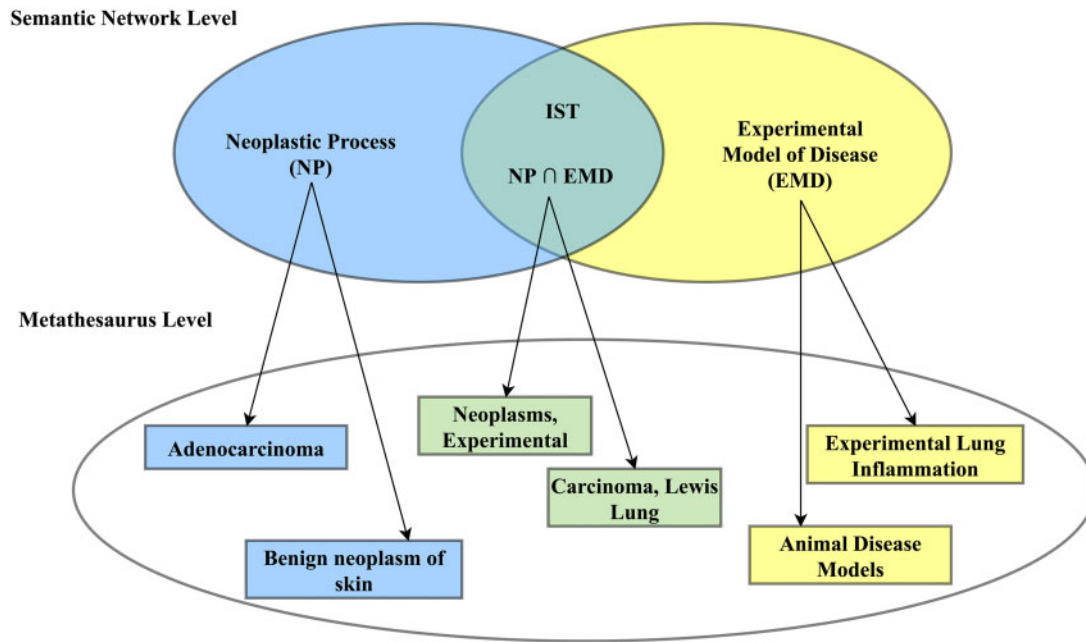


Figure 3. The 2 levels of the Unified Medical Language System. In the Semantic Network (SN) level, we have the semantic types *Neoplastic Process* (NP), *Experimental Model of Disease* (EMD), and the intersection semantic type (IST) $NP \cap EMD$ between them. The Metathesaurus level shows concepts assigned the intersection semantic type and the 2 pure semantic types, colored to correspond to the colors of their assigned semantic type. For example, the concept *Neoplasms, Experimental* (as suggested by its name) is assigned both STs.

gorithm overcoming the blockage was applied to the ST *Experimental Model of Disease* and found many extra concepts.

Geller et al⁷⁰ introduced 2 structural inconsistency patterns of the dual hierarchical relationships of 2 UMLS concepts and their STs: ST inversion (ie, ST assigned to child concept is more general than ST assigned to parent concept) and lack of ancestry. The former is a better indicator of errors than the latter. Wei et al¹⁹ evaluated ST assignment consistency in the UMLS using SNOMED CT Specimen concepts. Overlapping concepts in intersections of semantic uniformity groups, defined by concepts' structural features, are strong indicators of inconsistency. Gu et al²⁰ partitioned a SNOMED CT hierarchy into disjoint semantic uniformity groups based on concepts' STs in the UMLS. Concepts in small groups are more likely to have ST assignment errors. Gu et al⁷¹ detected ST assignment errors for UMLS concepts if their STs are inconsistent with the mapped STs of their SNOMED CT semantic tags. Mejino and Rosse⁷² illustrated that inconsistent ST assignments of anatomical concepts in UMLS can be reconciled following the principles of the Digital Anatomist⁷³ and Foundational Model of Anatomy.⁷⁴

Chen et al⁴⁷ modified RSN to model chemical concepts with multiple "Chemical Viewed Structurally" STs, identifying concepts with an invalid combination of STs and with incorrect ST assignments. Morrey et al⁷⁵ resolved redundant ST assignments for chemical composites with multiple STs based on the relative sizes of components in the molecular structure of a composite. A Chemical Specialty Semantic Network⁷⁶ was developed to provide a better categorization of chemical concepts in the UMLS. Rare STs in Chemical Specialty Semantic Network highlight errors.

Fan et al^{77,78} proposed a corpus-driven approach to auditing ST assignments. A huge set of >14 million MetaMap-processed PubMed abstracts was used to identify CUIs and their shallow-parsed contextual features. After CUI-to-ST grouping, a distribu-

tional classifier was trained for reclassifying CUIs into more appropriate semantic categories. Fan et al^{48,79,80} added a text classifier using the CUI lexical synonyms and found it complementary to the earlier context-based classifier.

He et al⁸¹ addressed auditing ST assignments for the 10 STs in the top levels of SN. By the specificity rule,⁶⁰ a concept is assigned the most specific ST possible. Hence, top STs should be assigned only to a few general or abstract concepts. Reviewers found that 2-thirds of these concepts have too general incorrect ST assignments. UMLS editors should avoid "erring up" in assigning top STs.

He et al²¹ monitored the longitudinal changes in ST assignments via the lens of the RSN.⁶¹ They showed that many intersections that were removed from RSN due to error reports reappear due to the categorization of new concepts to nonsensical or forbidden ST combinations. To cope with this problem, Geller et al²² created a rule-based system for the UMLS editors to test whether any combination of up to 5 STs is allowed, forbidden, or questionable.

Hierarchical (IS-A) relationships

Hierarchical relationships constitute the backbone of a terminology enabling inheritance of lateral relationships and enable efficient use of a whole class of concepts (eg, *Myocardial Infarction* and all its descendants) for information retrieval, data mining, etc. Researchers pay special attention to auditing the hierarchical relationships of the UMLS. In a terminology, a cycle of hierarchical relationships is forbidden. In the UMLS, such cycles indicate errors in or inconsistencies between sources. Detection and resolution of cycles between 2 and among 3 UMLS concepts are discussed by Bodenreider,³² Mouglin and Bodenreider,³³ and Halper et al⁶⁴ respectively. Pisanelli et al⁸² detected redundancies, cycles, and misuse of hierarchical relationships by an ontological analysis of the Metathesaurus.

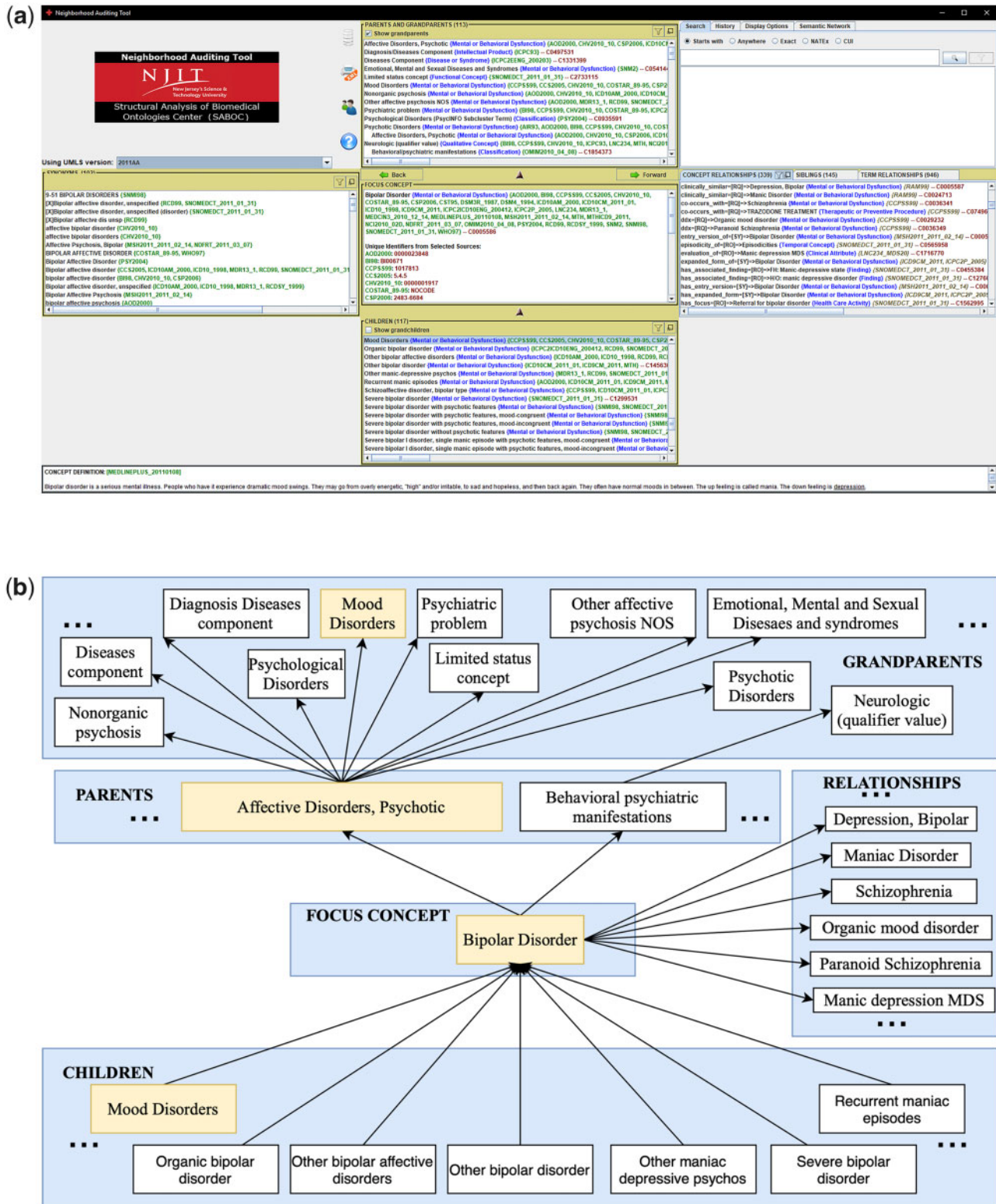


Figure 4. The Neighborhood Auditing Tool (NAT) interface and corresponding “neighborhood” network: (A) A screenshot of the NAT tool for the concept *Bipolar Disorder* (as in Figure 1 for the Unified Medical Language System [UMLS] interface): the focus concept is shown in the central box. The parents and grandparents in the top box (with indentation), and children (and grandchildren [not displayed]) in the bottom box. The synonyms are to the left and relationships (or siblings) are to the right. The semantic type for each concept in the screen is in blue, the UMLS sources in green, and the concept unique identifier in red. The number of concepts in each box overflows its capacity and the box is scrollable. This screenshot from 2011 is interesting because it is rich enough to display a forbidden cycle of 3 concepts.⁶⁴ *Mood Disorders* as the top child, → *Bipolar Disorders* as the focus concept, → *Affective Disorders, Psychotic* as a second parent (third line from the bottom), → *Mood Disorder* as the sixth grandparent, closing a cycle of 3 concepts. This error was reported to the UMLS team and this cycle does not exist in the UMLS 2019 AB version of *Bipolar Disorder* (Figure 1). (B) Excerpt of the neighborhood for *Bipolar Disorder*: the highlighted boxes in yellow shows the cycle of 3 concepts. The light blue rectangles correspond to the various windows in panel (A).

Bodenreider³⁴ examined redundancy and semantic consistency in hierarchical relationships in UMLS sources by indexing hierarchical paths between 2 concepts. A weak link is found between redundancy and semantic consistency. Semantic inconsistency in redundant hierarchical relationships indicates potential miscategorization. Xing et al⁸³ developed a tool (FEDRR) to detect redundant hierarchical relationships in source vocabularies, in linear time, using UMLS files. The overall completeness, consistency, and usability of the UMLS are evaluated by Bodenreider et al³⁵ using a multiaxial coding system (MAOUSSC). They note inconsistency in hierarchical relationships and a paucity of lateral relationships (year of study: 1998).

Bodenreider et al³⁶ further examined the occurrence of noun phrase modifiers for concepts to assess the consistency of biomedical terminologies. The study compared disease and procedure terms in SNOMED to the UMLS. They counted the frequency of modifier pairs (eg, acute and chronic, primary and secondary) in the noun phrases and noted the lack of certain terms and relationships. Another method, COHeRE (Cross-Ontology Hierarchical Relation Examination)⁸⁴ detects inconsistencies and possible errors in hierarchical relationships across UMLS sources. COHeRE leverages the UMLS knowledge sources and the MapReduce cloud computing technique for systematic, large-scale ontology QA. Research indicates the majority of inconsistent relationships exist in the sources rather than being introduced in the UMLS integration process.

An algorithm to identify missing IS-A relationships from concepts of an extent of a RST is described by Chen et al.⁸⁵ The extent of each RST is divided into singly rooted components. The recursive algorithm suggests for an editor to check missing IS-A relationships from the roots of small components to concepts of large components.

Gu et al²⁶ examined conflicting hierarchical relationships, redundant hierarchical relationships, mixed hierarchical and lateral relationships, and multiple lateral relationships in the UMLS. They investigated whether multiple relationships between 2 concepts are from the same source terminologies.

Lateral relationships

It is important to audit lateral relationships because they bear the nonhierarchical semantic connections between concepts. For certain concepts, it is impossible to reduce the ambiguity of relationships but possible to limit ambiguity by suggesting other relationship types. Mary et al⁸⁶ proposed to extend the relationships with several relationships defined between semantic types in SN, which may improve web searches. Other researchers¹⁷ focused on auditing concepts associated with multiple relationships that differ in granularity or are contradictory, heterogeneous, or homogeneous.

A semantic method was proposed¹⁸⁷ for auditing lateral relationships by transforming them into a relationship signature and mapping signatures from the Metathesaurus to the SN. Vizenor et al⁸⁷ argued that the semantics of lateral relationships need to be more explicitly defined by ontology developers and extend the SN.

Schulz and Hahn³⁸ created a terminological knowledge base using the Metathesaurus. They extracted anatomy and pathology concepts from the Metathesaurus and map them in a semi-automated way to a representation model that emphasizes part-whole reasoning. The process reveals inconsistencies in lateral relationships.

Topological pattern-based ontology enrichment

He et al²³ introduced a topological pattern-based ontology enrichment method for source ontologies in the UMLS. Topological pat-

terns are derived from the UMLS based on the IS-A links between identical pairs of concepts from 2 ontologies. An m:n pattern has m(n) IS-A links in the first (second) ontology. The intermediate concepts are candidates to enrich a source ontology pending the review of domain experts. This method can also help audit the UMLS by detecting missing synonyms or erroneous classifications. The 2:2, k:1, 1:k, and m:n patterns were considered to enrich SNOMED CT.^{23,24} In other studies, NCI Thesaurus (NCIt) was enriched.^{25,88} Additionally, a mathematical formula was used to compute the number of potential placements of new concepts in a target ontology.⁸⁹ The formula was extended to the cases where cross-ontology synonyms are possible.⁹⁰ These methods leveraged the vertical density differences between 2 ontologies. Keloth et al⁹¹ considered the horizontal density differences (number of siblings) in 2 ontologies. In most cases, the differences in sets of siblings are due to alternative classifications by ontology designers, not enabling enrichment. Keloth et al⁹² also used a mathematical criterion for likely cases of alternative classification to reduce human efforts for finding potential cases. They designed randomized controlled trials to compare the recommendations, with the decisions of a human expert.

Ontology alignment

Alignment techniques serve to investigate the equivalence between concepts based on various kinds of mapping across ontologies. The mappings are based on similar concept names, definitions, and relationships. Bodenreider and Burgun¹⁰ applied 2 methods based on lexical and conceptual similarity for node alignment of the SN with the UMLS Metathesaurus. Vizenor et al⁹³ aligned the Metathesaurus relationships with SN relationships. One of the main applications enabled by these alignment strategies is auditing the consistency between the SN and the Metathesaurus. A limited review uncovered wrong and missing occurrences in both ST assignments and hierarchical relationships. Schulz et al⁹⁴ provided methods and assessed the alignment of UMLS SN with BioTop^{95,96} and identified inconsistent multiple ST combinations.

Several studies have used UMLS synonymy to identify anchor concepts for point-to-point mappings across ontologies, such as NCIt, Adult Mouse Anatomical Dictionary,⁹⁷ MeSH,⁹⁸ ATC,⁹⁹ Foundational Model of Anatomy,⁷⁴ and GALEN,¹⁰⁰ which could not be found by lexical similarity and conceptual similarity.¹⁰¹⁻¹⁰⁷ When the inconsistencies uncovered in these studies of the different ontologies are corrected, they are indirectly updated in the new UMLS release. The results of the different alignment methods for anatomical ontologies were also compared and analyzed¹⁰⁸⁻¹¹⁰ along with the challenges of the alignment process.¹¹¹

Furthermore, ontology alignment and auditing have been facilitated by automated approaches, including logic-based and string similarity-based approaches. Jimenez-Ruiz et al¹¹²⁻¹¹⁴ described a mapping among NCIt, Foundational Model of Anatomy, and SNOMED CT utilizing UMLS as alignments reference. A logic-based semantics technique was illustrated to effectively detect errors in the UMLS, utilizing the conservativity, consistency and locality principles. To enhance source-integration and auditing, a SPED (Shortest Path Edit Distance) algorithm was proposed as a string similarity measure for UMLS terms.¹¹⁵

Summary of study properties

Table 3 depicts the categorization data of the reviewed articles. For each technique, we recorded the level of automation and the kind of knowledge source, 2 critical issues for applying the technique. This

3-dimensional table groups together studies with similar qualities, for example, those which report on auditing hierarchical relationships using automatic systematic techniques or which use only intrinsic knowledge. However, finding the automation level or knowledge type of a given article requires a search for its reference in some entries. Thus, Table 4 lists the qualities of each article. In Supplementary Table 1, we report on the auditing results for the articles. Figure 5 shows the distribution of articles of different categories over time.

DISCUSSION

Terminology developers variably follow desirable characteristics for terminology models.¹¹⁷ Errors and inconsistencies are therefore expected when integrating terminologies into the UMLS. The 83 reviewed studies demonstrate the special role UMLS plays in the field of QA of terminologies.^{13,116,118}

We have given a short description of the QA techniques in each of the 83 surveyed articles in order to describe the various available techniques in a single place. Classifying the techniques according to the aspects of a UMLS concept will enable practitioners like UMLS curators in the NLM and researchers to learn from previously developed techniques, say for QA of synonyms or hierarchical relationships.

From Table 3, we observe:

1. Most studies ($n = 55$) combine the use of intrinsic and extrinsic knowledge (IEK) sources.
2. Most studies ($n = 56$) use automated heuristic (AH) techniques. Some studies ($n = 26$) are automated systematic (AS), and very few ($n = 3$) are purely manual (MN), while 46 are both AH and IEK.
3. The 2 most common aspects are concepts, concept names, and synonyms ($n = 37$) and ST assignments ($n = 36$).

As much as researchers try to develop automatic techniques, there is typically a need for a domain expert review of the results. This is not surprising since QA is as complicated as terminology modeling, which is not automatic. Only specific errors, which are detectable by logical rules, can be totally automated (eg, redundancy in ST assignment, hierarchical, and lateral relationships).

That the 2 most investigated aspects are concepts, concept names, synonyms and ST assignments, which may be because they are the most important features of the UMLS and are widely used for various downstream applications including natural language processing, data mining, information retrieval, mapping from local terminologies, creation of clinical data warehouses, etc. In addition, errors in these aspects can be corrected in the UMLS itself, as explained in Introduction. The techniques concentrated on ST assignment utilize the mapping from the Metathesaurus to the SN, which is an Abstraction Network designed independently for the UMLS to capture the semantics of concepts,⁶¹ in contrast to Abstraction Networks which are derived from a terminology.^{119–121} Thus, the mapping of concepts to STs provides a reality check on the mapping of terms to concepts. In particular, several studies demonstrate that concepts assigned to multiple STs are susceptible to errors due to their semantic complexity. Designers of terminologies could mimic the UMLS in creating an a priori Abstraction Network for a terminology rather than an a posteriori Abstraction Network as has been done previously.⁶¹ Learning from the UMLS experience, such networks can help with terminology QA. Finally, an error in

the ST assignment might indicate confusion or ambiguity about the semantics of a concept which may be manifested in the existence of other errors.

This is only one way that the unique design of the UMLS opened research opportunities in the field of auditing biomedical terminologies. As a compendium knowledge base that integrates multiple terminologies, UMLS opens the possibility of comparing and contrasting multiple terminologies. This quality was exploited in the research on alignments of terminologies and topological-based ontology enrichment.

In the Materials and Methods section, we described how some kinds of errors occur in the UMLS. Errors migrating to the UMLS from source terminologies can be detected by the context in which the modeling of several terminologies is contrasted. These errors were not detected in the context of their own terminologies. But even errors made in the UMLS, like erroneously matching concepts from different terminologies to the same UMLS concept, provide valuable feedback on semantic issues and improper naming to the source terminologies curators.

Figure 5 shows a constant flow of approximately 3 articles a year. The peaks in 2007 and 2009 show the impact funding can have on this niche research area. In some subjects one can trace research progress along a particular approach. For example, auditing IS-A cycles of length 2 started with Bodenreider³² and continued with Mougin and Bodenreider,³³ and extended for length 3 in Halper et al.⁶⁴

Another example of longitudinal progress is found regarding likelihood of errors in small ISTs. The early detection of errors was a side effect.^{40,41} Several rigorous studies established this observation.^{43,44} In Gu et al,⁶² all small ISTs concepts were reviewed. In Gu et al,^{45,71} 2 refinements were presented. In Chen et al⁴⁷ and Morrey et al,^{75,76} special modeling was required for chemical concepts and ISTs in which intersections are frequent. Then, in He et al,²¹ a longitudinal study showed that, while NLM corrected reported errors, eliminating nonsensical or forbidden ISTs (by SN use notes), those ISTs pop up again after a year or 2. To provide a systematic solution preventing such cases, Geller et al²² designed a system to check the legitimacy of any IST before assigning it to concepts. This chain of articles demonstrates a development from an accidental observation, through studies reported to the UMLS, to the creation of a tool to provide a systematic solution to prevent errors by editors, obtaining better quality, and saving resources.

The topological pattern-based method started with vertical density differences between a pair of source terminologies in terms of IS-A paths.^{48–50} It continued to construct different topological patterns.^{23–25,88–90} Recently, Keloth et al.^{91,92} expanded to investigate the horizontal density differences between source terminologies.

Alignment research began with specific ontologies.^{101–103,111} The techniques evolved from manual to automated, rule-based systems and hybrid strategies combining direct and indirect alignment techniques,^{104,106} with Zhang and Bodenreider^{109,110} summarizing lessons learned.

Future research

To predict future directions for auditing the Metathesaurus, we take cues from recent trends in QA of UMLS source terminologies (eg, SNOMED, GO, NCI).

1. Initial efforts harness machine learning (ML) for QA of hierarchical (IS-A) relationships.^{122–125} A critical issue for ML is

Table 3. A 3-dimensional table categorizing studies by the aspects audited, the automation level and the kind of knowledge used

Automation level	Knowledge source		
	Intrinsic knowledge (n = 23)	Extrinsic knowledge (n = 5)	Intrinsic and extrinsic knowledge (n = 55)
Concepts, concept names, and synonyms (37 references)			
Automated systematic (n = 7)	63,65	30,57	31,52,115
Automated heuristic (n = 29)	11,27–29,55,56		23–25,36,88–92,101–114
Manual (n = 1)			58
Semantic type assignment (36 references)			
Automated systematic (n = 13)	11,21,27,46,63,65,70	77,78	22,48,79,80
Automated heuristic (n = 21)	87		10,18–20,35,37,39–41,43–45,47,62,68,69,71,75,76,94
Manual (n = 2)		81	72
Hierarchical relationships (24 references)			
Automated systematic (n = 13)	11,26,27,32,33,42,63,64,65,82–84		31
Automated heuristic (n = 11)	17,34,38,55,56		10,35,36,85,86,93
Lateral relationships (12 references)			
Automated systematic (n = 5)	11,26,63,65		31
Automated heuristic (n = 7)	17,38,87		35,36,86,93

Table 4. Direct access for categorization data

Ref	ASPE	AT	KNW	Ref	ASPE	AT	KNW	Ref	ASPE	AT	KNW
10	STA	AH	IEK	38	HREL	AH	IK	75	STA	AH	IEK
	HREL				LREL						
11	CCNS	AHASASAS	IK	39	STA	AH	IEK	76	STA	AH	IEK
	STA		IK								
	HREL		IK								
	REL		IK								
17	HREL	AH	IK	40	STA	AH	IEK	77,78	STA	AS	EK
	LREL										
18	STA	AH	IEK	41	STA	AH	IEK	48,79,80	STA	AS	IEK
19	STA	AH	IEK	42	HREL	AS	IK	81	STA	MN	EK
20	STA	AH	IEK	43	STA	AH	IEK	82	HREL	AS	IK
21	STA	AS	IK	44	STA	AH	IEK	83	HREL	AS	IK
22	STA	AS	IEK	45	STA	AH	IEK	84	HREL	AS	IK
23	CCNS	AH	IEK	46	STA	AS	IK	85	HREL	AH	IEK
24	CCNS	AH	IEK	47	STA	AH	IEK	86	HRELLREL	AH	IEK
25	CCNS	AH	IEK	52	CCNS	AS	IEK	87	LRELSTA	AH	IK
26	HREL	AS	IK	55,56	CCNS	AH	IK	88	CCNS	AH	IEK
	LREL				HREL						
27	STA	AS	IK	57	CCNS	AS	EK	89	CCNS	AH	IEK
	CCNS	AH	IK								
	HREL	AS	IK								
28	CCNS	AH	IK	58	CCNS	MN	IEK	90	CCNS	AH	IEK
29	CCNS	AH	IK	62	STA	AH	IEK	91	CCNS	AH	IEK
30	CCNS	AS	EK	63	ALL	AS	IK	92	CCNS	AH	IEK
31	HREL	AS	IEK	64	HREL	AS	IK	93	HREL	AH	IEK
	LREL								LREL		
	CCNS										
32	HREL	AS	IK	65	ALL	AS	IK	94	STA	AH	IEK
33	HREL	AS	IK	68	STA	AH	IEK	101–107	CCNS	AH	IEK
34	HREL	AH	IK	69	STA	AH	IEK	108–111	CCNS	AH	IEK
35	HREL	AH	IEK	70	STA	AS	IK	112–114	CCNS	AH	IEK
	LREL										
	STA										
36	HREL	AH	IEK	71	STA	AH	IEK	115	CCNS	AS	IEK
	LREL										
	CCNS										
37	STA	AH	IEK	72	STA	MN	IEK				

This table enables direct access to the categorization properties for each study.

AH: automated heuristic; AS: automated systematic; ASPE: aspect; AT: automation level; CCNS: concepts, concept names, and synonyms; EK: extrinsic knowledge; HREL: hierarchical relationships; IEK: intrinsic and extrinsic knowledge; IK: intrinsic knowledge; KNW: knowledge source; LREL: lateral relationships; MN: manual; Ref: reference; STA: semantic type assignment.

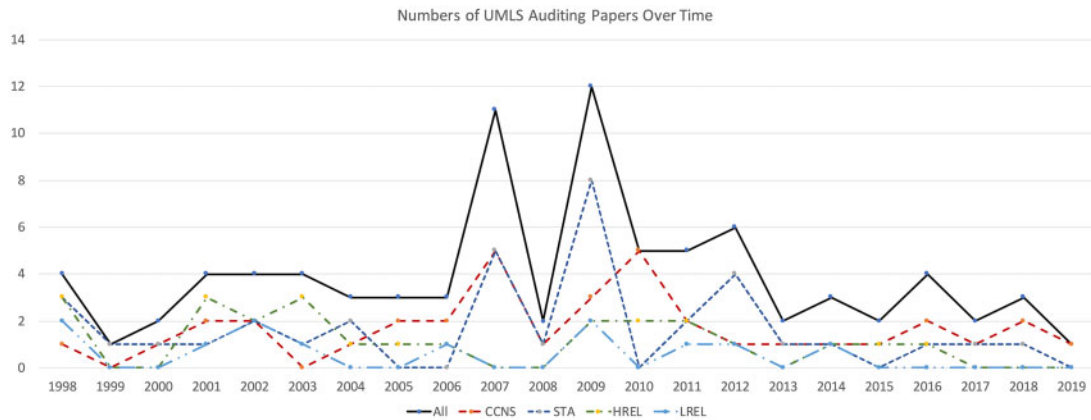


Figure 5. Publication trend over time. The trends of the numbers of publications about Unified Medical Language System (UMLS) auditing between 1998 and 2019, stratified by different aspects of a UMLS concept. Note that an article may audit several aspects of a concept so the total may be less than the sum of all the aspects. Overall, there are 2 surges of publications in 2007 and 2009 with 11 and 12 articles, respectively, possibly due to National Library of Medicine funding support on UMLS quality assurance 2005-2009 and the first special issue on terminology auditing in 2009.¹³ Except for those 2 years, there were on average about 3 publications a year. During 2010-2012, there are still more late publications due to above funding. In the last 7 years we see a decline of interest in quality assurance of the UMLS, with an average of 2.4 articles per year. For example, the second special issue on terminology auditing¹¹⁶ in 2018 did not include any UMLS articles. In 2007, most articles were focused on concept names and synonyms and semantic type assignments (STAs). In 2009, most articles were about auditing STAs, while in 2010, most articles were focused on concept names and synonyms. The numbers of articles that audited relationships were consistently low, but there were more articles on auditing hierarchical relationships (HREL) than lateral relationships (LREL). CCNS: concepts, concept names, and synonyms.

obtaining training data, perhaps by comparing consecutive releases of UMLS tracing error corrections.

2. The techniques we reviewed are based on one idea. In the terminology QA, we observe a recent trend of hybrid techniques combining multiple ideas. Hybrid techniques hold promise for better performance. For example, Cui et al¹²⁶ combined the structural technique of nonlattice with the natural language processing technique. Others combine structural and lexical techniques.¹²⁷⁻¹²⁹ Two ML techniques mentioned previously^{123,125} are hybrid.
3. Scalable approaches based on distributed computing framework for big data (eg, MapReduce) audit IS-A relationships in UMLS⁸⁴ and SNOMED CT.¹³⁰⁻¹³³

To reduce the level of human efforts involved in auditing, we expect that ML, hybrid, and big data techniques will improve the auditing yield by increasing the ratio of errors found to the number of concepts reviewed in the UMLS and reducing false positives. Improvements in this area will lead to greater application of the methods.

CONCLUSION

The UMLS contains innumerable errors and inconsistencies of varying importance that originate in its component terminologies or in the addition of its unique features. Researchers of QA techniques for the UMLS have found creative methods to expose errors and inconsistencies in this enormous problem space. This exhaustive survey of the state of the art in UMLS auditing will assist researchers, terminology resource developers, and advanced UMLS users to identify and adapt existing methods that may be applicable to their own needs.

FUNDING

JC is supported in part by research funds from the University of Alabama School of Medicine Informatics Institute and by the Center for Clinical and

Translational Sciences under grant UL1TR001417 and the National Center for Advancing Translational Sciences. ZH is supported in part by the University of Florida Clinical and Translational Science Institute funded by National Center for Advancing Translational Sciences under award number UL1TR001427 and National Institute on Aging awards R01AG064529 and R21AG061431. LL is supported by Medical Informatics Fellowship with Veteran Administration funding from the Office of Academic Affairs, Department of Veterans Affairs. LZ is supported by Monmouth University Summer Faculty Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

YP, JC, and XZ conceived, designed, guided, and coordinated the study and the writing. ZH and DW identified publication records from MEDLINE and screened the titles, abstracts, and full text of the articles. ZL identified additional articles that were not retrieved in MEDLINE. VK prepared the figures. The tables were prepared by LZ, VK, DW, and ZH. Each of the 9 authors reviewed and summarized a subset of articles. All the authors contributed to the writing of the article. The Abstract, Introduction, Discussion, and Conclusion were written jointly by YP and JC. ZH wrote the Materials and Methods section. LL performed thorough editing of the article. All the authors revised and approved the final article.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their insightful remarks and suggestions, which helped significantly to improve this article.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

- Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004; 32 (90001): 267D.
- Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. *Proc AMIA Symp* 1998; 1998: 815–9.
- Humphreys BL, Lindberg DA, Hole WT. Assessing and enhancing the value of the UMLS Knowledge Sources. *Proc Annu Symp Comput Appl Med Care* 1991; 1991: 78–82.
- Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med* 1993; 32 (4): 281–91.
- Humphreys BL, Lindberg DA. Building the unified medical language system. *Proc Annu Symp Comput Appl Med Care* 1989; 1989: 475–80.
- Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc* 1998; 5 (1): 1–11.
- McCray AT. The UMLS semantic network. *Proc Annu Symp Comput Appl Med Care* 1989; 1989: 503–7.
- McCray AT, Hole WT. The scope and structure of the first version of the UMLS semantic network. *Proc Annu Symp Comput Appl Med Care* 1990; 1990: 126–30.
- McCray AT. An upper-level ontology for the biomedical domain. *Int J Genomics* 2003; 4 (1): 80–4.
- Bodenreider O, Burgun A. Aligning knowledge sources in the UMLS: methods, quantitative results, and applications. *Stud Health Technol Inform* 2004; 107 (Pt 1): 327–31.
- Cimino JJ. Auditing the Unified Medical Language System with semantic methods. *J Am Med Inform Assoc* 1998; 5 (1): 41–51.
- Chen Y, Perl Y, Geller J, Cimino JJ. Analysis of a study of the users, uses, and future agenda of the UMLS. *J Am Med Inform Assoc* 2007; 14 (2): 221–31.
- Geller J, Perl Y, Halper M, Cornet R. Special issue on auditing of terminologies. *J Biomed Inform* 2009; 42 (3): 407–11.
- Zhu X, Fan J-W, Baorto DM, Weng C, Cimino JJ. A review of auditing methods applied to the content of controlled biomedical terminologies. *J Biomed Inform* 2009; 42 (3): 413–25.
- Amith M, He Z, Bian J, Lössio-Ventura JA, Tao C. Assessing the practice of biomedical ontology evaluation: Gaps and opportunities. *J Biomed Inform* 2018; 80: 1–13.
- Maedche A, Staab S. Measuring similarity between ontologies In: Gómez-Pérez A, Benjamins VR, eds. *International Conference on Knowledge Engineering and Knowledge Management*. New York, NY: Springer; 2002: 251–63.
- Mougin F, Grabar N. Auditing the multiply-related concepts within the UMLS. *J Am Med Inform Assoc* 2014; 21 (e2): e185–93.
- Mougin F, Bodenreider O, Burgun A. Analyzing polysemous concepts from a clinical perspective: application to auditing concept categorization in the UMLS. *J Biomed Inform* 2009; 42 (3): 440–51.
- Wei D, Halper M, Elhanan G. Using SNOMED semantic concept groupings to enhance semantic-type assignment consistency in the UMLS. In: *proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. 2012: 825–30.
- Gu H, Chen Y, He Z, Halper M, Chen L. Quality assurance of UMLS semantic type assignments using SNOMED CT hierarchies. *Methods Inf Med* 2016; 55 (2): 158–65.
- He Z, Morrey CP, Perl Y, et al. Sculpting the UMLS refined semantic network. *Online J Public Health Inform* 2014; 6 (2): e181.
- Geller J, He Z, Perl Y, Morrey CP, Xu J. Rule-based support system for multiple UMLS semantic type assignments. *J Biomed Inform* 2013; 46 (1): 97–110.
- He Z, Geller J, Elhanan G. Categorizing the relationships between structurally congruent concepts from pairs of terminologies for semantic harmonization. *AMIA Jt Summits Transl Sci Proc* 2014; 2014: 48–53.
- He Z, Geller J, Chen Y. A comparative analysis of the density of the SNOMED CT conceptual content for semantic harmonization. *Artif Intell Med* 2015; 64 (1): 29–40.
- He Z, Chen Y, de Coronado S, Piskorski K, Geller J. Topological-pattern-based recommendation of UMLS concepts for National Cancer Institute thesaurus. *AMIA Annu Symp Proc* 2016: 618–27.
- Gu H, Elhanan G, Halper M, He Z. Questionable relationship triples in the UMLS. In: *proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*. 2012: 713–6.
- Cimino JJ. Battling Scylla and Charybdis: the search for redundancy and ambiguity in the 2001 UMLS metathesaurus. *Proc AMIA Symp* 2001; 2001: 120–4.
- Hole WT, Srinivasan S. Discovering missed synonymy in a large concept-oriented Metathesaurus. *Proc AMIA Symp* 2000; 2000: 354–8.
- Huang KC, Geller J, Halper M, Cimino JJ. Piecewise synonyms for enhanced UMLS source terminology integration. *Proc AMIA Symp* 2007; 2007: 339–43.
- Huang KC, Geller J, Halper M, Perl Y, Xu J. Using WordNet synonym substitution to enhance UMLS source integration. *Artif Intell Med* 2009; 46 (2): 97–109.
- Bodenreider O, Mitchell JA, McCray AT. Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. *Proc AMIA Symp* 2002; 2002: 61–5.
- Bodenreider O. Circular hierarchical relationships in the UMLS: etiology, diagnosis, treatment, complications and prevention. *Proc AMIA Symp* 2001; 2001: 57–61.
- Mougin F, Bodenreider O. Approaches to eliminating cycles in the UMLS Metathesaurus: naive vs. formal. In *Proc AMIA Symp* 2005; 2005: 550–4.
- Bodenreider O. Strength in numbers: exploring redundancy in hierarchical relations across biomedical terminologies. *Proc AMIA Symp* 2003; 2003: 101–5.
- Bodenreider O, Burgun A, Botti G, Fieschi M, Le Beux P, Kohler F. Evaluation of the Unified Medical Language System as a medical knowledge source. *J Am Med Inform Assoc* 1998; 5 (1): 76–87.
- Bodenreider O, Burgun A, Rindfleisch TC. Assessing the consistency of a biomedical terminology through lexical knowledge. *Int J Med Inform* 2002; 67 (1–3): 85–95.
- Chen Y, Gu HH, Perl Y, Geller J, Halper M. Structural group auditing of a UMLS semantic type's extent. *J Biomed Inform* 2009; 42 (1): 41–52.
- Schulz S, Hahn U. Medical knowledge reengineering—converting major portions of the UMLS into a terminological knowledge base. *Int J Med Inform* 2001; 64 (2–3): 207–21.
- Geller J, Gu H, Perl Y, Halper M. Semantic refinement and error correction in large terminological knowledge bases. *Data Knowl Eng* 2003; 45 (1): 1–32.
- Gu H, Perl Y, Geller J, Halper M, Liu LM, Cimino JJ. Modeling the UMLS using an OODB. *Proc AMIA Symp* 1999; 1999: 82–6.
- Gu H, Perl Y, Geller J, Halper M, Liu LM, Cimino JJ. Representing the UMLS as an object-oriented database: modeling issues and advantages. *J Am Med Inform Assoc* 2000; 7 (1): 66–80.
- Cimino JJ, Min H, Perl Y. Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus. *J Biomed Inform* 2003; 36 (6): 450–61.
- Gu H, Perl Y, Elhanan G, Min H, Zhang L, Peng Y. Auditing concept categorizations in the UMLS. *Artif Intell Med* 2004; 31 (1): 29–44.
- Gu HH, Hripesak G, Chen Y, Morrey CP, et al. Evaluation of a UMLS auditing process of semantic type assignments. *AMIA Annu Symp Proc* 2007; 2007: 294–8.
- Gu HH, Min H, Peng Y, Zhang L, Perl Y. Using the metaschema to audit UMLS classification errors. *AMIA Annu Symp Proc* 2002; 2002: 310–4.
- Peng Y, Halper MH, Perl Y, Geller J. Auditing the UMLS for redundant classifications. *AMIA Annu Symp Proc* 2002; 2002: 612–6.

47. Chen L, Morrey CP, Gu H, Halper M, Perl Y. Modeling multi-typed structurally viewed chemicals with the UMLS Refined Semantic Network. *J Am Med Inform Assoc* 2009; 16 (1): 116–31.
48. Fan J-W, Xu H, Friedman C. Using contextual and lexical features to re-structure and validate the classification of biomedical concepts. *BMC Bioinformatics* 2007; 8 (1): 264.
49. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; 339: b2535.
50. Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. *Proc AMIA Symp* 2001; 2001: 662–6.
51. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; 25 (1): 25–9.
52. Huang K-C, Geller J, Elhanan G, Perl Y, Halper M. Auditing SNOMED Integration into the UMLS for duplicate concepts. *Proc AMIA Symp* 2010; 2010: 321–5.
53. Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. *J Biomed Inform* 2003; 36 (6): 414–32.
54. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. *Stud Health Technol Inform* 2001; 84 (Pt 1): 216–20.
55. Erdogan H, Bodenreider O, Erdem E. Finding semantic inconsistencies in UMLS using answer set programming. In: proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence; 2010: 1927–8.
56. Erdogan H, Erdem E, Bodenreider O. Exploiting UMLS semantics for checking semantic consistency among UMLS concepts. *Stud Health Technol Inform* 2010; 160 (Pt 1): 749–53.
57. Liu H, Lussier YA, Friedman C. A study of abbreviations in the UMLS. *Proc AMIA Symp* 2001; 2001: 393–7.
58. Merrill GH. Concepts and synonymy in the UMLS metathesaurus. *J Biomed Discov Collab* 2009; 4: 7.
59. McCray AT. Representing biomedical knowledge in the UMLS semantic network. In: Broering NC, ed. *High Performance Medical Libraries: Advances in Information Management for the Virtual Era*. New York, NY: Meckler; 1993: 45–55.
60. McCray AT, Nelson SJ. The representation of meaning in the UMLS. *Methods Inf Med* 1995; 34 (1/2): 193–201.
61. Halper M, Gu H, Perl Y, Ochs C. Abstraction networks for terminologies: Supporting management of big knowledge. *Artif Intell Med* 2015; 64 (1): 1–16.
62. Gu HH, Elhanan G, Perl Y, et al. A study of terminology auditors' performance for UMLS semantic type assignments. *J Biomed Inform* 2012; 45 (6): 1042–8.
63. Morrey CP, Geller J, Halper M, Perl Y. The Neighborhood Auditing Tool: a hybrid interface for auditing the UMLS. *J Biomed Inform* 2009; 42 (3): 468–89.
64. Halper M, Morrey CP, Chen Y, Elhanan G, Hripcsak G, Perl Y. Auditing hierarchical cycles to locate other inconsistencies in the UMLS. *Proc AMIA Symp* 2011; 2011: 529–36.
65. Ochs C, Geller J, Perl Y. A relationship-centric hybrid interface for browsing and auditing the UMLS. *J Integr Des Process Sci* 2011; 15 (4): 3–25.
66. Halper MH, Chen Z, Geller J, Perl Y. A metaschema of the UMLS based on a partition of its semantic network. *Proc AMIA Symp* 2001; 2001: 234–8.
67. Perl Y, Chen Z, Halper M, Geller J, Zhang L, Peng Y. The cohesive metaschema: a higher-level abstraction of the UMLS semantic network. *J Biomed Inform* 2002; 35 (3): 194–212.
68. Chen Y, Gu H, Perl Y, Halper M, Xu J. Expanding the extent of a UMLS semantic type via group neighborhood auditing. *J Am Med Inform Assoc* 2009; 16 (5): 746–57.
69. Chen Y, Gu H, Perl Y, Geller J. Overcoming an obstacle in expanding a UMLS semantic type extent. *J Biomed Inform* 2012; 45 (1): 61–70.
70. Geller J, Morrey CP, Xu J, et al. Comparing inconsistent relationship configurations indicating UMLS errors. *Proc AMIA Symp* 2009; 2009: 193–7.
71. Gu H, He Z, Wei D, Elhanan G, Chen Y. Validating UMLS semantic type assignments using SNOMED CT semantic tags. *Methods Inf Med* 2018; 57 (1/2): 43–53.
72. Mejino J Jr, Rosse C. The potential of the digital anatomist foundational model for assuring consistency in UMLS sources. *Proc AMIA Symp* 1998; 1998: 825–9.
73. Rosse C, Shapiro LG, Brinkley JF. The digital anatomist foundational model: principles for defining and structuring its concept domain. *Proc AMIA Symp* 1998; 1998: 820–4.
74. Rosse C, Mejino JL. The foundational model of anatomy ontology. In: Burger A, Davidson D, Baldoock R, eds. *Anatomy Ontologies for Bioinformatics*. Berlin, Germany: Springer; 2008: 59–117.
75. Morrey CP, Chen L, Halper M, Perl Y. Resolution of redundant semantic type assignments for organic chemicals in the UMLS. *Artif Intell Med* 2011; 52 (3): 141–51.
76. Morrey CP, Perl Y, Halper M, Chen L, Gu HH. A chemical specialty semantic network for the Unified Medical Language System. *J Cheminform* 2012; 4 (1): 9.
77. Fan JW, Friedman C. Semantic classification of biomedical concepts using distributional similarity. *J Am Med Inform Assoc* 2007; 14 (4): 467–77.
78. Fan JW, Xu H, Friedman C. Using distributional analysis to semantically classify UMLS concepts. *Stud Health Technol Inform* 2007; 129 (Pt 1): 519–23.
79. Fan JW, Friedman C. Combining contextual and lexical features to classify UMLS concepts. *Proc AMIA Symp* 2007; 2007: 231–5.
80. Fan J-W, Friedman C. Semantic reclassification of the UMLS concepts. *Bioinformatics* 2008; 24 (17): 1971–3.
81. He Z, Perl Y, Elhanan G, Chen Y, Geller J, Bian J. Auditing the assignments of top-level semantic types in the UMLS semantic network to UMLS concepts. *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 2017; 2017: 1262–9.
82. Pisanelli DM, Gangemi A, Steve G. An ontological analysis of the UMLS Metathesaurus. *Proc AMIA Symp* 1998; 1998: 810–4.
83. Xing G, Zhang GQ, Cui L. FEDRR: fast, exhaustive detection of redundant hierarchical relations for quality improvement of large biomedical ontologies. *BioData Min* 2016; 9: 31.
84. Cui L. COHeRE: cross-ontology hierarchical relation examination for ontology quality assurance. *Proc AMIA Symp* 2015; 2015: 456–65.
85. Chen Y, Gu HH, Perl Y, Geller J. Structural group-based auditing of missing hierarchical relationships in UMLS. *J Biomed Inform* 2009; 42 (3): 452–67.
86. Mary V, Le Duff F, Mouglin F, Le Beux P. Method for automatic management of the semantic network ambiguity in the UMLS: possible application for information retrieval on the Web. *Stud Health Technol Inform* 2003; 95: 475–9.
87. Vizenor LT, Bodenreider O, McCray AT. Auditing associative relations across two knowledge sources. *J Biomed Inform* 2009; 42 (3): 426–39.
88. He Z, Chen Y, Geller J. Perceiving the usefulness of the National Cancer Institute metathesaurus for enriching NCIt with topological patterns. *Stud Health Technol Inform* 2017; 245: 863–7.
89. He Z, Geller J. Preliminary analysis of difficulty of importing pattern-based concepts into the National Cancer Institute Thesaurus. *Stud Health Technol Inform* 2016; 228: 389–93.
90. He Z, Keloth VK, Chen Y, Geller J. Extended analysis of topological-pattern-based ontology enrichment. *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 2018; 2018: 1641–8.
91. Keloth VK, He Z, Chen Y, Geller J. Leveraging horizontal density differences between ontologies to identify missing child concepts: a proof of concept. *Proc AMIA Symp* 2018; 2018: 644–53.
92. Keloth VK, He Z, Elhanan G, Geller J. Alternative classification of identical concepts in different terminologies: different ways to view the world. *J Biomed Inform* 2019; 94: 103193.
93. Vizenor L, Bodenreider O, Peters L, McCray AT. Enhancing biomedical ontologies through alignment of semantic relationships: exploratory approaches. *Proc AMIA Symp* 2006; 2006: 804–8.

94. Schulz S, Beisswanger E, van den Hoek L, Bodenreider O, van Mulligen EM. Alignment of the UMLS semantic network with BioTop: methodology and assessment. *Bioinformatics* 2009; 25 (12): i69–76.
95. Beisswanger E, Schulz S, Stenzhorn H, Hahn U. BioTop: An upper domain ontology for the life sciences. *Appl Ontol* 2008; 3 (4): 205–12.
96. Schulz S, Beisswanger E, Hahn U, Wermter J, Kumar A, Stenzhorn H. From GENIA to BIOTOP: towards a top-level ontology for biology. In: proceedings of the 2006 Conference on Formal Ontology in Information Systems (FOIS 2006); 2006: 103–14.
97. Hayamizu TF, Mangan M, Corradi JP, Kadin JA, Ringwald M. The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biol* 2005; 6 (3): R29.
98. Nelson SJ, Schopen M, Savage AG, Schulman JL, Arluk N. The MeSH translation maintenance system: structure, interface design, and implementation. *Stud Health Technol Inform* 2004; 107 (Pt 1): 67–9.
99. WHO Collaborating Centre for Drug Statistics Methodology. ATC structure and principles. https://www.whocc.no/atc/structure_and_principles/ Accessed June 23, 2020.
100. Rector AL, Nowlan WA, Consortium G. The GALEN project. *Comput Methods Prog Biomed* 1994; 45 (1–2): 75–8.
101. Bodenreider O, Hayamizu TF, Ringwald M, De Coronado S, Zhang S. Of mice and men: aligning mouse and human anatomies. *Proc AMIA Symp* 2005; 2005: 61–5.
102. Winnenburger R, Rodriguez L, Callaghan FM, Sorbello A, Szarfman A, Bodenreider O. Aligning pharmacologic classes between MeSH and ATC. In: proceedings of the 4th International Conference on Biomedical Ontology (ICBO); 2013.
103. Zhang S, Bodenreider O. Alignment of multiple ontologies of anatomy: Deriving indirect mappings from direct mappings to a reference. *Proc AMIA Symp* 2005; 2005: 864–8.
104. Zhang S, Bodenreider O. NLM anatomical ontology alignment system results of the 2006 ontology alignment contest. In: proceedings of the 1st International Conference on Ontology Matching-Volume 225; 2006: 153–64.
105. Zhang S, Bodenreider O. Aligning multiple anatomical ontologies through a reference. In proceedings of the 1st International Conference on Ontology Matching-Volume 225; 2006: 201–5.
106. Zhang S, Bodenreider O. Hybrid alignment strategy for anatomical ontologies results of the 2007 ontology alignment contest. In: proceedings of the 2nd International Conference on Ontology Matching-Volume 304; 2007: 139–49.
107. Zhang S, Mork P, Bodenreider O, Bernstein PA. Comparing 2 approaches for aligning representations of anatomy. *Artif Intell Med* 2007; 39 (3): 227–36.
108. Zhang S, Mork P, Bodenreider O. Lessons learned from aligning two representations of anatomy. In: proceedings of the First International Workshop on Formal Biomedical Knowledge Representation (KR-MED); 2004: 102–8.
109. Zhang S, Bodenreider O. Experience in aligning anatomical ontologies. *Int J Semant Web Inf Syst* 2007; 3 (2): 1–26.
110. Zhang S, Bodenreider O. Lessons learned from cross-validating alignments between large anatomical ontologies. *Stud Health Technol Inform* 2007; 129 (Pt 1): 822–6.
111. Bodenreider O. Issues in mapping LOINC laboratory tests to SNOMED CT. *AMIA Annu Symp Proc* 2008; 2008: 51–5.
112. Jiménez-Ruiz E, Grau BC, Horrocks I. Exploiting the UMLS Metathesaurus in the Ontology Alignment Evaluation Initiative. In: proceedings of the 2nd International Workshop on Exploiting Large Knowledge Repositories; 2012.
113. Jiménez-Ruiz E, Grau BC, Horrocks I, Berlanga R. Logic-based assessment of the compatibility of UMLS ontology sources. *J Biomed Semantics* 2011; 2 (Suppl 1): S2.
114. Jimenez-Ruiz E, Grau BC, Horrocks I, Llavori RB. Towards a UMLS-based silver standard for matching biomedical ontologies. In: proceedings of the 5th International Workshop on Ontology Matching (OM-2010); 2010.
115. Rudnyi A, Geller J, Song M. Shortest Path Edit Distance for Enhancing UMLS Integration and Audit. *AMIA Annu Symp Proc* 2010; 2010: 697–701.
116. Geller J, Perl Y, Cui L, Zhang GQ. Quality assurance of biomedical terminologies and ontologies. *J Biomed Inform* 2018; 86: 106–8.
117. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med* 1998; 37 (4/5): 394–403.
118. Agrawal A, Cui L, eds. 1st International Workshop on Quality Assurance of Biological and Biomedical Ontologies and Terminologies. <https://home.mannhattan.edu/~ankur.agrawal/workshop/bibm2018> Accessed June 23, 2020.
119. Gu H, Halper M, Geller J, Perl Y. Benefits of an object-oriented database representation for controlled medical terminologies. *J Am Med Inform Assoc* 1999; 6 (4): 283–303.
120. Min H, Perl Y, Chen Y, Halper M, Geller J, Wang Y. Auditing as part of the terminology design life cycle. *J Am Med Inform Assoc* 2006; 13 (6): 676–90.
121. Wang Y, Halper M, Min H, Perl Y, Chen Y, Spackman KA. Structural methodologies for auditing SNOMED. *J Biomed Inform* 2007; 40 (5): 561–81.
122. Liu H, Geller J, Halper M, Perl Y. Using convolutional neural networks to support insertion of new concepts into SNOMED CT. *AMIA Annu Symp Proc* 2018; 2018: 750–9.
123. Abeyesinghe R, Qu X, Cui L. Identifying similar non-lattice subgraphs in gene ontology based on structural isomorphism and semantic similarity of concept labels. *AMIA Annu Symp Proc* 2018; 2018: 1186–95.
124. Liu H, Perl Y, Geller J. Transfer learning from BERT to support insertion of new concepts into SNOMED CT. *AMIA Annu Symp Proc* 2019; 2019: 1129–38.
125. Zheng L, Liu H, Perl Y, Geller J. Training a convolutional neural network with terminology summarization data improves SNOMED CT enrichment. *AMIA Annu Symp Proc* 2019; 2019: 972–81.
126. Cui L, Zhu W, Tao S, Case JT, Bodenreider O, Zhang GQ. Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT. *J Am Med Inform Assoc* 2017; 24 (4): 788–98.
127. Agrawal A, Perl Y, Ochs C, Elhanan G. Algorithmic detection of inconsistent modeling among SNOMED CT concepts by combining lexical and structural indicators. *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 2015; 2015: 476–83.
128. Agrawal A, Perl Y, Chen Y, Elhanan G, Liu M. Identifying Inconsistencies in SNOMED CT problem lists using structural indicators. *AMIA Annu Symp Proc* 2013; 2013: 17–26.
129. Agrawal A, Elhanan G. Contrasting lexical similarity and formal definitions in SNOMED CT: consistency and implications. *J Biomed Inform* 2014; 47: 192–8.
130. Zhang GQ, Zhu W, Sun M, Tao S, Bodenreider O, Cui L. MaPLE: A MapReduce pipeline for lattice-based evaluation and its application to SNOMED CT. In: proceedings of the IEEE International Conference on Big Data; 2014: 754–9.
131. Cui L, Tao S, Zhang G-Q. Biomedical ontology quality assurance using a big data approach. *ACM Trans Knowl Discov Data* 2016; 10 (4): 1–28.
132. Tao S, Cui L, Zhu W, Sun M, Bodenreider O, Zhang G-Q. Mining relation reversals in the evolution of SNOMED CT using MapReduce. *AMIA Jt Summits Transl Sci Proc* 2015; 2015: 46–50.
133. Zhu W, Zhang G-Q, Tao S, Sun M, Cui L. NEO: systematic non-lattice embedding of ontologies for comparing the subsumption relationship in SNOMED CT and in FMA using MapReduce. *AMIA Jt Summits Transl Sci Proc* 2015; 2015: 216–20.