

Auditing National Cancer Institute thesaurus neoplasm concepts in groups of high error concentration

Ling Zheng^{a,*}, Hua Min^b, Yan Chen^c, Julia Xu^d, James Geller^a and Yehoshua Perl^a

^a Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA

^b Department of Health Administration and Policy, College of Health and Human Services, George Mason University, Fairfax, VA 22030-4444, USA

^c CIS Department, Borough of Manhattan Community College, CUNY, New York, NY 10007, USA

^d National Library of Medicine, Bethesda, MD 20894, USA

Abstract. The National Cancer Institute thesaurus is an important knowledge resource that should ideally be error-free. We investigated the occurrence of errors in the *Neoplasm* subhierarchy, which is a part of the National Cancer Institute thesaurus *Disease, Disorder or Finding* hierarchy. There are five key findings in this study. (1) Errors in the *Neoplasm* subhierarchy are not uniformly distributed. (2) A *partial-area taxonomy*, which is a compact network for summarizing the structure and content of an ontology, helped uncover groups of concepts, called “small partial-areas,” in the *Neoplasm* subhierarchy. (3) The rate of errors in “small partial-areas” is twice as large as in “large partial-areas” (44% versus 22%), satisfying statistical significance. Thus, we conclude that higher error concentrations exist in small partial-areas. (4) *Group-based auditing* can be used successfully to identify additional suspicious concepts in a small group, once a few members of the group are already known as erroneous. (5) *Error correction propagation* can be used successfully and with minimal effort to correct additional errors in the *Neoplasm* subhierarchy that occur outside of an initial small group of erroneous concepts. We present examples of errors and examples of how corrections transform and simplify the partial-area taxonomy.

Keywords: NCI thesaurus, ontology quality assurance, ontology auditing, abstraction network, error concentration

Accepted by: Barry Smith

1. Introduction

The National Cancer Institute thesaurus (NCIt) is commonly used in healthcare information systems and cancer research. Examples include decision support systems, interoperability of information systems, data annotation, cancer phenotyping and disease database integration (Bodenreider, 2008; Campbell et al., 2015; Hochheiser et al., 2016; Jiang et al., 2015; Park et al., 2016; Rubin et al., 2008). Thus it is important to perform quality assurance (QA) to guarantee the accuracy of NCIt’s content to prevent modeling errors to propagate into application systems. However, the scarcity of QA resources, such as trained domain expert time, makes it next to impossible to perform a comprehensive QA project on the whole ontology or even on one of the 19 top-level hierarchies of it (e.g., on the *Disease, Disorder or Finding* hierarchy of NCIt, which contains 25,360 concepts). Even its important *Neoplasm* subhierarchy, which is critical for the NCI mission, consists of 8,166 concepts.

*Corresponding author. Tel.: +1 (862) 371-0988; Fax: +1 (973) 596-5236; E-mail: lz265@njit.edu.

There are two major activities that lead to corrections of ontologies. (1) Curators of ontologies receive occasional requests of users to correct modeling errors they find, but such requests are *ad hoc* and do not constitute a rigorous QA process. (2) Ontology maintenance teams execute internal QA processes to test and verify the correctness of every new release of an ontology. See, for example, the internal QA process of NCIt in place at the National Cancer Institute (NCI) as described by De Coronado et al. (2009). Automated QA processes can only expose errors detectable by algorithms, such as redundant role assignments. Such QA algorithms can detect structural errors but not semantic errors, which are more difficult to uncover.

Hence, there is a need for a rigorous QA process as an integral part of the life cycle of an ontology that detects semantic errors as well (Min et al., 2006). As in finance, software verification, etc., such QA processes should not be the responsibility of the editorial team of an ontology, but be outsourced to an external department or even an external organization that has no emotional attachment to the modeling decisions of the ontology and thus can be objective in an ontology review. In business software such an outside auditing group is often referred to as “the red team.”

Considering the fact that ontology errors are created as a result of unintentional human mistakes, rather than occurring as natural phenomena, one might think that they will be distributed uniformly over the concepts of an ontology. However, in this study we refute the assumption that errors are uniformly distributed in the investigated medical ontology. While this phenomenon is not so obvious when viewing an ontology with existing visualization tools that do not perform summarization, it becomes clear when viewing the same ontology through the prism of an Abstraction Network, which, as described below, provides guidance for where to look for errors. Therefore, an economical approach to the QA of ontologies is to identify structural characterizations of sets of concepts for which a relatively high rate of errors is expected, compared to a random control sample. Reviewing such sets of concepts by domain experts is expected to provide a high QA yield, measured by the ratio of concepts confirmed as erroneous for a given number of reviewed concepts.

How can we identify such sets of “suspicious” concepts? If errors are localized, how can we identify the parts of an ontology with high concentrations of errors? This paper will offer an answer to this question, based on *Abstraction Networks*. In a long-range research program conducted by the Structural Analysis of Biomedical Ontologies Center (SABOC; <https://sites.google.com/a/njit.edu/saboc/home>), algorithmic methods for deriving Abstraction Networks from ontologies have been developed (Halper et al., 2015). An Abstraction Network is a compact summary network consisting of “nodes,” each representing a set of concepts that are similar in their structure and semantics. Thus, the Abstraction Network summarizes the structure and content of an ontology. One specific kind of Abstraction Network is called *partial-area taxonomy* (Min et al., 2006; Ochs et al., 2016; Wang et al., 2007). Partial-area taxonomies have been used to summarize hierarchies of large ontologies such as NCIt (De Coronado et al., 2004), SNOMED CT (Stearns et al., 2001) and Gene Ontology (GO; Ashburner et al., 2000). They have also been used to summarize hierarchies of small OWL ontologies (He et al., 2013a, 2013b; Ochs et al., 2012, 2013). These small ontologies include the Cancer Chemoprevention Ontology (CanCo; Zeginis et al., 2014), the Drug Discovery Investigations Ontology (DDI; Qi et al., 2010), the Ontology of Clinical Research (OCRe; Sim et al., 2014), and the Sleep Domain Ontology (SDO; Arabandi et al., 2010).

The compact view of an ontology offered by the partial-area taxonomy of the ontology offers ways to identify sets of concepts with higher rates of errors than for the rest of the ontology. In other words, in such cases, our methodologies successfully direct QA experts to locations with higher error concentrations, as demonstrated here for the *Neoplasm* subhierarchy of NCIt. Furthermore, Wei & Bodenreider

(2010) have shown that using partial-area taxonomies enables finding errors that are not detectable by automatic classifiers.

We show also that the use of taxonomic *group-based auditing* as well as taxonomic *error correction propagation*, methods newly developed for this paper, will be demonstrated to further spread error identification and correction, increasing the QA yield. Moreover these corrections lead to removal of some suspicious sets of concepts from the partial-area taxonomy, by incorporating their concepts into other (non-suspicious) sets of concepts.

2. Background

2.1. National Cancer Institute thesaurus

The National Cancer Institute thesaurus (NCIt) has played an important role in biomedical research through various applications. For example, the Community Clinical Oncology Program (CCOP), a clinical trial program, uses NCIt to code patient care information (“EVS Use and Collaborations,” 2014). NCIt was utilized to build a cancer drug dictionary for drug name normalization in drug datasets (Jiang et al., 2015). Donfack et al. (2012) developed an ontology-driven decision support system for medical diagnosis based on NCIt. Besides, NCIt has been adopted by many organizations as a framework to develop terminology standards. The FDA has chosen NCIt to develop and publish important terminology sets, for example, drug submissions, and device event problems (Reed & Kaufman-Rivi, 2010).

NCIt is a biomedical ontology, modeled using description logic, and published monthly in various formats (e.g., in OWL format) by the National Cancer Institute. The content of NCIt covers a broad range of domains including clinical care, basic research, public information and administrative activities to facilitate cancer translational research. NCIt has been applied in various biomedical information systems within and outside of NCI for data sharing and semantic interoperability (Fragoso et al., 2004).

The NCI thesaurus exists in two versions, the asserted and the inferred version. The asserted version contains assertions explicitly defined by the NCIt team, while the inferred version is obtained by running a reasoner on the asserted version. In the version we used for this research (February 2015 inferred version in OWL format), there are 108,376 active concepts, which are organized into 19 IS-A hierarchies such as *Disease, Disorder or Finding*; *Biological Process*; *Gene*; *Drug, Food, Chemical or Biomedical Material* and *Anatomic Structure, System, or Substance*.

A concept may have multiple parents, but can only exist in one of the 19 top-level hierarchies of NCIt. Concepts are defined by roles and inherit roles from their ancestors along the hierarchy. Each role has its associated domain and range and the name of each role contains its (abbreviated) domain and range names. For example, the role *Disease Has Associated Anatomic Site* has as domain the *Disease, Disorder or Finding* hierarchy and as range the *Anatomic Structure, System, or Substance* hierarchy where the specific pathological process is located.

Out of the 19 hierarchies, eight hierarchies such as the *Abnormal Cell* hierarchy and the *Conceptual Entity* hierarchy have no roles. They serve as ranges for roles from other hierarchies. Each of the other 11 hierarchies has a list of associated permitted role types. For example, the *Disease, Disorder or Finding* hierarchy has 29 role types, including *Disease Excludes Abnormal Cell* with the range *Abnormal Cell* hierarchy, *Disease Has Finding* with the range *Disease, Disorder or Finding* hierarchy, *Disease Mapped To Gene* with the range *Gene* hierarchy, and *Disease Has Normal Cell Origin* with the range *Anatomic Structure, System, or Substance* hierarchy.

The extension of NCIt is driven by user requests, and cancer-related concepts are modeled with a higher priority and more detail than are other concepts. In the February 2015 release, the *Disease, Disorder or Finding* hierarchy, which is the largest hierarchy in NCIt, contained 25,360 concepts ($23.4\% = 25,360/108,376$) with 7.79 roles on average. In this research, we focused on the *Neoplasm* subhierarchy with 8,166 concepts ($32.2\% = 8,166/25,360$) in the *Disease, Disorder or Finding* hierarchy, with an average of 23.02 roles.

The NCIt team has applied various automated and manual QA techniques to assure the quality of NCIt during its whole life cycle (De Coronado et al., 2009). Many external QA reviews were also conducted for NCIt. For example, Ceusters et al. (2005) performed a terminological and ontological analysis of NCIt. For the *Biological Process* hierarchy and the *Gene* hierarchy of NCIt different kinds of errors were found using Abstraction Networks (Min et al., 2006, 2008). Mougin & Bodenreider (2008) applied Semantic Web technologies to audit NCIt. Zhu et al. (2009) assembled an extensive review of QA techniques for NCIt and medical terminologies in general, which appeared in a journal special issue on QA of terminologies (Geller et al., 2009).

2.2. Partial-area taxonomy

In our previous research, we have designed different kinds of Abstraction Networks to summarize ontologies and to support quality assurance (QA) for large ontologies including NCIt, SNOMED CT and Gene Ontology (GO) (Halper et al., 2015). In general, an Abstraction Network is a compact summary network of an ontology that preserves much of its structure and semantics. The two kinds of Abstraction Networks most commonly used are called *area taxonomy* and *partial-area taxonomy* (Min et al., 2006; Wang et al., 2007).

In this study, we derived the *partial-area taxonomy* of a part of NCIt according to the roles used as class restrictions by subclass axioms, called *restriction-defined taxonomies* (Ochs et al., 2013). Partial-areas are sets of concepts with similar structure and similar semantics. In a partial-area taxonomy all such similar concepts are represented by a single node. Nodes are connected by *child-of* links that are derived from the IS-A links of the ontology. The diagram of a partial-area taxonomy is always considerably smaller than the diagram of the original ontology. Partial-areas are not necessarily disjoint, due to concepts with multiple parents/ancestors. For more details on how to construct a partial-area taxonomy and for examples of how to use partial-area taxonomies see the work of Min et al. (2006) and of Wang et al. (2007).

2.3. Meta-methodology for ontology auditing: Looking for uncommon combinations of structure and semantics

In this subsection, we formulate and explain the meta-methodology, which looks for “uncommonly” modeled concepts, viewing them as concepts suspected of modeling errors. By modeling error we refer to wrong or missing features in the description of a concept (e.g., definition, synonyms, hierarchical relationships (IS-A) and semantic relationships (roles)) with respect to the existing biomedical knowledge. In addition, ontological rules should be followed. An example of an ontological rule is that roles are inherited down along the IS-A hierarchy. This meta-methodology will be clarified for its manifestation in small partial-areas, in the context of partial-area taxonomies.

A node in a partial-area taxonomy (as defined in Section 2.2) represents a particular set of concepts of the ontology. These concepts are similar in their structure and semantics. That is, they all share the same roles and the same root. When encountering the partial-area *Neoplasm* with 403 concepts, the modeling of its concepts is considered “common” – there are many concepts with the same neoplasm semantics and structural modeling – with the role *Disease Has Abnormal Cell*. However, when we encounter a small partial-area of, say, two concepts only, and their combination of structure and semantics is unique to them among the thousands of concepts in the hierarchy, then this would be a case of “uncommon modeling,” since no other concepts have the same combination of structure and semantics.

It is, of course, possible that an ontology correctly contains only two concepts that are represented with a specific structure and semantics. However, another option is that the reason for this uncommon modeling is that there is an error in how these two concepts are represented in the ontology. If so, once this error is corrected, say by adding a role or changing a parent link, then these concepts are likely to reappear in another partial-area to reflect the changes in their structures or semantics. It may well be the case that the new “home partial-area” is not small. This was an example scenario where the modeling was “uncommon” due to an error, and correcting the error(s) eliminated a small partial-area.

In work of Halper et al. (2007), small partial-areas were indeed found to be characterized by higher error rates. More precisely, higher error rates were observed for partial-areas of up to seven concepts in the (relatively small) *Specimen* hierarchy of SNOMED CT. Min et al. (2006) also obtained a similar result for the small *Biological Process* hierarchy of NCIt, but this time for partial-areas of up to three concepts. In neither of these studies was statistical significance established, but they showed higher error rates for the concepts from small partial-areas in comparison to larger partial-areas.

In both these studies, the threshold x between small and large partial-areas was chosen in a way that all partial-areas up to and including size x were considered small and all partial-areas with sizes larger than x were considered large. The value of x was not predetermined, but selected to maximize the difference between the error rate for the small and the large partial-areas. The option of medium size partial-areas was not considered. More differences between the current study and previous studies are listed in the beginning of the Discussion Section.

Interestingly, the same meta-methodology, with a different manifestation, played a major role in the research of the SABOC team on QA techniques for the Unified Medical Language System (UMLS) (Bodenreider, 2004; Humphreys et al., 1998), the flagship terminological system of the National Library of Medicine. One part of the UMLS is the Semantic Network (McCray, 2003; McCray & Nelson, 1995).

In our research, we showed that when a specific combination of multiple semantic types of the UMLS Semantic Network is assigned to only a very small number of concepts, then there is a high likelihood of errors (Geller et al., 2013; Gu et al., 2004, 2007, 2012; He et al., 2014; Morrey et al., 2009, 2012). In this case, the identification of these kinds of sets of concepts was achieved by the creation of the Refined Semantic Network (RSN) (Gu et al., 2000, 2007; He et al., 2014). Thus, we view the UMLS Semantic Network as well as our Refined Semantic Network as Abstraction Networks for the UMLS. The existence of small sets of concepts with a unique combination of semantic types was the manifestation of uncommon modeling, which exposed errors in those studies. This corresponds to the uncommon modeling of small partial-areas used in this study. In both cases, the “prism” constituted by Abstraction Networks was needed for the detection of sets with uncommon modeling, which were found to contain relatively higher ratios of errors.

3. Methods

3.1. Materials and approach

The *Neoplasm* subhierarchy of NCIt with 8,166 concepts is of special importance because of the priority given to modeling cancer-related concepts due to the mission of NCIt to support cancer research and care. Thus, the NCIt team is paying increased attention to the modeling of neoplasm concepts compared to many other concepts in NCIt.

A proof of the high priority given to the *Neoplasm* subhierarchy can be found, in a quantifiable way, by comparing the density of the *Neoplasm* subhierarchy to the density of the remainder of the *Disease, Disorder or Finding* hierarchy. Although there are more non-neoplasm concepts (17,194) in the *Disease, Disorder or Finding* hierarchy, their total number of roles is smaller. The reason is that 14,336 non-neoplasm concepts ($83.4\% = 14,336/17,194$) have no roles at all. The hierarchical density, measured by the average number of parents, is 1.73 for the *Neoplasm* subhierarchy versus 1.10 for the non-neoplasm concepts. The lateral density, measured by the average number of instances of roles for a concept, is 23.02 for the *Neoplasm subhierarchy*, versus 0.55 for all non-neoplasm concepts. Limiting ourselves to concepts that have *any roles*, the average number of roles for the 2,858 non-neoplasm concepts is 3.33.

Therefore, we approached the problem whether we can identify a location where there is a higher concentration of erroneous concepts in the relatively large *Neoplasm* subhierarchy. Can we confirm, with statistical significance, the following hypothesis?

Hypothesis 1. *Small partial-areas in the Neoplasm subtaxonomy of the Disease, Disorder or Finding hierarchy of NCIt harbor sets of concepts with higher rates of errors than large partial-areas.*

Having the partial-area taxonomy at our disposal, we randomly selected 150 concepts from small partial-areas, defined here as having sizes between 1 and 10, as study sample. For the control sample, 40 concepts were randomly selected from large partial-areas of at least 20 concepts. The range 11–19 is considered to define medium-sized partial-areas. The study sample and the control sample were combined and the order of concepts was randomized.

The review of the selected concepts involved two steps following the modified Delphi procedure (Dalkey & Helmer, 1963). Namely, first, each of three domain experts, being blind to the sampling technique that was used, independently reviewed all the concepts in the sample and reported all erroneous concepts. The three error reports were combined into a questionnaire, where each error identified by any of the experts was listed, without attributing it to the expert reviewer(s) who discovered it. For every reported error a suggested correction was included.

In the second step of the process, each of the reviewers marked whether she agreed with each error in the combined report or not. A concept is considered a “consensus erroneous concept” only if all three reviewers agreed that this concept has an error. Finally, we compared the numbers and percentages of “consensus erroneous concepts” in small partial-areas and in large partial-areas with each other. The QA analysis on the 190 neoplasm concepts was carried out by three of the co-authors (YC, HM and JX) who are domain experts in medicine with extensive experience in ontology QA.

3.2. Example of group-based auditing

The partial-area taxonomy of an ontology divides concepts into groups. The basic idea of the “group-based” auditing method is that if “many” concepts in a group are found to exhibit errors, then a conscientious auditor should review all the concepts of that group (Wang et al., 2012). This is supported by

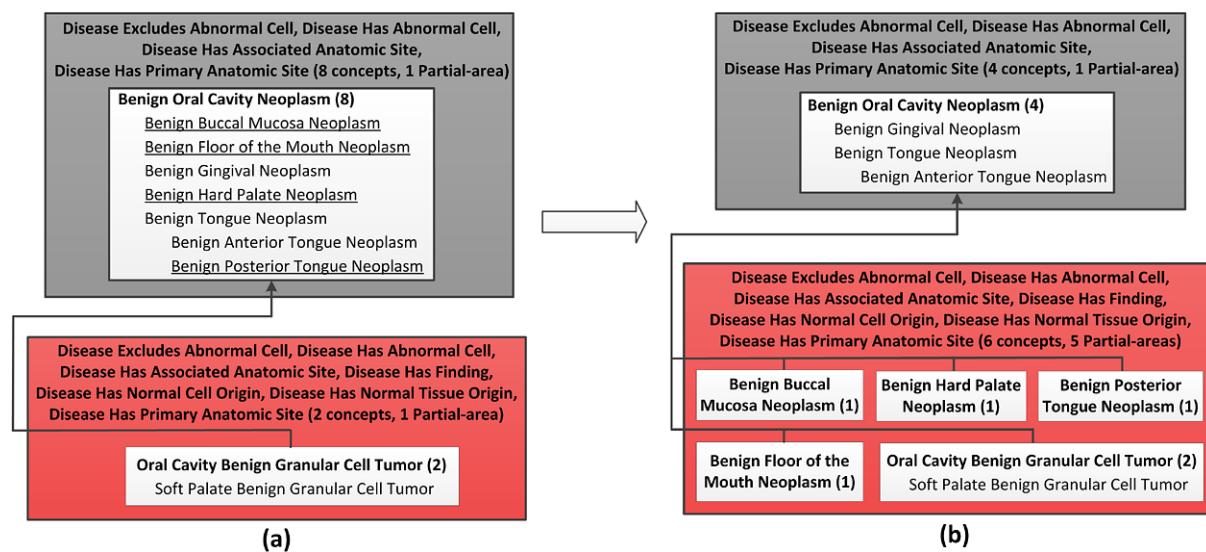


Fig. 1. Example of the structure of the *Neoplasm* partial-area taxonomy (a) before and (b) after auditing.

work of He et al. (2014), which showed that it is advisable to review the other concepts of a group if some errors were found in the group. This is the abstract idea of group-based auditing.

Now follows a concrete example for illustrating group-based auditing, utilizing the partial-area taxonomy where the partial-areas function as groups. Figure 1 demonstrates a group-based auditing scenario. White boxes within colored boxes are partial area nodes within area nodes. The indented format in each partial-area in Fig. 1(a) represents the *IS-A* hierarchical structure inside of a partial-area node (e.g., *Benign Posterior Tongue Neoplasm IS-A Benign Tongue Neoplasm*, which *IS-A Benign Oral Cavity Neoplasm*). This detailed information is normally *not* shown in a partial-area taxonomy diagram, but is necessary for the demonstration of group-based auditing. The concept in bold (e.g., *Benign Oral Cavity Neoplasm*) in each partial-area node is the root, and the arrows denote the hierarchical *child-of* links between partial-area nodes. For example, the root concept *Oral Cavity Benign Granular Cell Tumor IS-A Benign Oral Cavity Neoplasm*, so there is a *child-of* link from the partial-area node *Oral Cavity Benign Granular Cell Tumor (2)* to the partial-area node *Benign Oral Cavity Neoplasm (8)*. The number in parentheses () is the number of concepts summarized by a node.

Out of the eight concepts in the partial-area node *Benign Oral Cavity Neoplasm (8)* in the top area node (Fig. 1(a)), only the four underlined concepts were in the random sample sent to our auditors (rows 1, 2, 4 and 7 under *Benign Oral Cavity Neoplasm (8)*). The auditors recommended adding the following three roles to these four concepts to improve the correctness of the modeling: *Disease Has Finding* with the target *Benign Cellular Infiltrate*, *Disease Has Normal Cell Origin* with the target *Connective and Soft Tissue Cell* and *Disease Has Normal Tissue Origin* with the target *Connective and Soft Tissue*. As noted above, this was a consensus decision.

Due to the suggested corrections, these four concepts should not remain in their current location. Rather they should appear in the lower area node {*Disease Excludes Abnormal Cell, Disease Has Abnormal Cell, Disease Has Associated Anatomic Site, Disease Has Finding, Disease Has Normal Cell Origin, Disease Has Normal Tissue Origin, Disease Has Primary Anatomic Site*}. When the partial-area taxonomy is re-derived, as shown in Fig. 1(b), it is indeed the case that each of the four concepts now

8

L. Zheng et al. / Auditing National Cancer Institute thesaurus neoplasm concepts in groups



Fig. 2. Example of error correction propagation and the resultant partial-area taxonomy simplification; (a) shows the partial-area taxonomy before and (b) after the auditing/correction steps.

appears in the lower area node. Furthermore, each of the four concepts is a root in the area node, thus it gets its own partial-area.

After observing the above errors of concepts in our sample, we raised the question whether the other four concepts in the same partial-area node might have the same errors as the four concepts in Fig. 1(a). In other words, we applied group-based auditing to the remaining four concepts in the area node at the top of Fig. 2(a). These concepts were not in the random sample that was originally given to the auditors.

The consensus auditing result of the three domain experts confirmed that the four concepts *Benign Oral Cavity Neoplasm*, *Benign Gingival Neoplasm*, *Benign Tongue Neoplasm* and *Benign Anterior Tongue Neoplasm* were also missing the same three roles as the four concepts that were in the original sample. Thus, group-based auditing in this case doubles the number of errors found, with little extra effort.

3.3. Error correction propagation and partial-area taxonomy simplification

Once correction of errors is achieved for a whole group, potentially through group-based auditing, one should consider the propagation of the correction of errors to descendant groups. Errors of a parent group are typically inherited by descendant groups. Thus, one can easily examine the inheritance of the corrections suggested for such errors.

We will now demonstrate the method of *error correction propagation*, where the corrections of the above errors, discovered for concepts within a partial-area **A**, will be propagated to three descendant partial-areas of the partial-area **A**. This happens as follows.

During group-based auditing we determined that the root concept *Benign Oral Cavity Neoplasm*, describing the semantics of the whole partial-area, is missing the above three roles. Because of that, the 22 concepts in Fig. 2(a) should all have the same roles, due to inheritance of the corrected set of relationships.

The concepts in the partial-area nodes of the area nodes in second and third level from the top in Fig. 2(a) originally had some, but not all, of the three missing roles discovered in Fig. 1. However, those missing roles are now added due to inheritance from the corrected *Benign Oral Cavity Neoplasm* root concept. After adding these three roles, the 22 concepts in the eight partial-area nodes in Fig. 2(a) will be summarized by only *one* partial-area node *Benign Oral Cavity Neoplasm* (22) in Fig. 2(b), which is not small anymore.

Figure 2 demonstrates an interesting visual impact of the error correction propagation process described above, namely partial-area taxonomy simplification. The eight partial-area nodes in four different area nodes appearing in Fig. 2(a) are unified into one single partial-area node of 22 concepts in Fig. 2(b). In Section 2.3, we discussed the idea that some small partial-areas are not reflecting the wide variety of concepts in the domain, but are the results of modeling errors. That is, modeling errors are manifested when more small partial-areas “than needed” appear. The process described by Fig. 2 shows the flip side of this observation, namely that the correction of modeling errors may lead to a beneficial simplification of the partial-area taxonomy (the summary) of the ontology, by unifying several small partial-area nodes into one larger partial-area node. Furthermore, due to inheritance, the simplification occurred across several area nodes. Interestingly, a similar simplification of the Refined Semantic Network, due to correction of errors, was shown by He et al. (2014), due to the disappearance of “small intersection semantic types” after reassigning their concepts to other semantic types.

4. Results

The *Neoplasm* subhierarchy, containing 8,166 concepts (32.25% of the whole *Disease, Disorder or Finding* hierarchy) in the February 2015 release of NCIt, is summarized by 4,824 partial-area nodes in the *Neoplasm* partial-area taxonomy. The three experts (see Method section) found 76 concepts out of 190 ($40.0\% = 76/190$) having errors. Table 1 shows the number of audited concepts (that appeared in the random sample), the number of erroneous concepts and the error rate for each partial-area node size. For example, as explained in Section 3.2, the random sample contained 10 concepts from partial-area nodes with size = 1, and among these the experts found five concepts (50.0%) exhibiting errors.

According to the results in Table 1, we can derive that the average error rate for the concepts from small partial-area nodes with up to 10 concepts is 44.7% ($= 67/150$), while the error rate for the concepts from large partial-area nodes, summarizing at least 20 concepts, is 22.5%. Table 2 summarizes the comparison

10

L. Zheng et al. / Auditing National Cancer Institute thesaurus neoplasm concepts in groups

Table 1

Distribution of erroneous concepts according to partial-area node size in the partial-area taxonomy

Partial-area node size	# of Concepts Audited	# of Erroneous Concepts	Error Rate (%)
1	10	5	50.0%
2	17	3	17.6%
3	10	5	50.0%
4	15	7	46.7%
5	13	3	23.1%
6	12	6	50.0%
7	21	8	38.1%
8	17	9	52.9%
9	19	10	52.6%
10	16	11	68.8%
11–19	–	–	–
≥ 20	40	9	22.5%
Total:	190	76	40.0%

Table 2

The 2×2 contingency table for small partial-areas and large partial-areas

	# Erroneous Concepts	# Concepts w/o Errors	Error Rate
Small partial-area nodes (size < 11)	67	83	44.7%
Large partial-area nodes (size ≥ 20)	9	31	22.5%

Table 3

Comparison of error distribution by types between concepts from small and large partial-area nodes

Error Type	# Erroneous Concepts from Small Partial-area nodes	# Erroneous Concepts from Large Partial-area nodes	Total
Missing role	60	5	65
Missing parent	13	3	16
Missing child	1	1	2
Incorrect parent	0	1	1

of these two error rates. The error rate difference is statistically significant ($p < 0.05$ according to Fisher's exact test; Good, 2005). Therefore, the results confirm Hypothesis 1, namely that concepts represented by small partial-area nodes are more likely to have errors than concepts in large partial-area nodes.

With one exception, all errors reported by our experts are errors of omission. Table 3 summarizes the distribution of erroneous concepts according to different error types. The most common type of error is the omission of roles, which occurs for 60 concepts from small partial-area nodes ($40.0\% = 60/150$) and five concepts from large partial-area nodes ($12.5\% = 5/40$). The second most common type of error is "missing parent" with 13 ($8.67\% = 13/150$) concepts having this error among small partial-area nodes and three ($7.5\% = 3/40$) among large partial-area nodes.

Table 4 lists some examples of concepts having modeling errors in their IS-A relationships. For example, the concept *Benign Epithelial Neoplasm* from a partial-area node summarizing only one concept is missing an IS-A relationship to *Benign Neoplasm*. Another concept *Benign Iris Neoplasm* from a small partial-area node was found missing a child *Iris Nevus*. The experts also found one concept, *Combined*

Table 4
Examples of erroneous hierarchical relationships

Partial-area node size	Error Type	Example Concept with Error	Reported Error
small	Missing parent	<i>Benign Epithelial Neoplasm</i>	Missing parent <i>Benign Neoplasm</i>
small	Missing child	<i>Benign Iris Neoplasm</i>	Missing child <i>Iris Nevus</i>
large	Missing parent	<i>Reproductive Endocrine Neoplasm</i>	Missing parent <i>Endocrine Neoplasm</i>
large	Missing child	<i>Intraventricular Brain Neoplasm</i>	Missing child <i>Glioblastoma</i> and <i>Pilocytic Astrocytoma</i>
large	Incorrect parent	<i>Combined Carcinoid and Adenocarcinoma</i>	Change the parent from <i>Carcinoma</i> to <i>Adenocarcinoma</i>

Table 5
The number of concepts from small partial-area nodes missing roles for each role type

Role Type	# Erroneous Concepts	Example Concept Missing Role Type	Target of Missing Role
<i>Disease Has Finding</i>	32	<i>Benign Cerebellar Neoplasm</i>	<i>Benign Cellular Infiltrate</i>
<i>Disease Excludes Abnormal Cell</i>	10	<i>Acantholytic Squamous Cell Skin Carcinoma</i>	<i>Malignant Basaloid Cell</i>
<i>Disease Excludes Finding</i>	8	<i>Amelanotic Melanoma</i>	<i>Favorable Clinical Outcome</i>
<i>Disease Has Primary Anatomic Site</i>	7	<i>High Grade Vaginal Intraepithelial Neoplasia</i>	<i>Vagina</i>
<i>Disease Has Normal Tissue Origin</i>	6	<i>Anal Canal Neuroendocrine Tumor</i>	<i>Neuroendocrine Tissue</i>
<i>Disease Has Normal Cell Origin</i>	5	<i>Granulosa Cell Tumor</i>	<i>Granulosa Cell</i>
<i>Disease Excludes Primary Anatomic Site</i>	4	<i>Acute Erythroid Leukemia</i>	<i>Lymphatic System</i>
<i>Disease Has Associated Anatomic Site</i>	3	<i>Hemolymphangioma</i>	<i>Lymphatic Vessel</i>
<i>Disease Excludes Normal Tissue Origin</i>	2	<i>Acute Erythroid Leukemia</i>	<i>Lymphoid Tissue</i>
<i>Disease Has Associated Disease</i>	2	<i>Adenocarcinoma in Multiple Adenomatous Polyps</i>	<i>Polypsis</i>
<i>Disease Has Abnormal Cell</i>	1	<i>Clear Cell Adenoma</i>	<i>Neoplastic Clear Cell</i>
<i>Disease Mapped To Gene</i>	1	<i>Borderline Ovarian Clear Cell Adenofibroma</i>	<i>ARID1A Gene</i>

Carcinoid and Adenocarcinoma, from a large partial-area node having an incorrect IS-A relationship.

For each hierarchy in NCIt, there is a list of role types that define the concepts in this hierarchy. The *Disease, Disorder or Finding* hierarchy has 29 role types and the name of each role type begins with “Disease” (e.g., *Disease Excludes Abnormal Cell*). For the 150 concepts from small partial-area nodes, our experts found 60 concepts missing 12 role types. Table 5 shows the distribution of these 60 erroneous concepts by role types and gives an example of one erroneous concept for each role type. For example, there are 10 concepts from small partial-area nodes missing the role *Disease Excludes Abnormal Cell*. *Acantholytic Squamous Cell Skin Carcinoma* is one of the 10 erroneous concepts, and it is missing this role with the target *Malignant Basaloid Cell*. Another issue discovered was that 32 concepts are missing the *Disease Has Finding* role.

The error correction propagation method (Section 3.4) was applicable to a number of concepts in our sample, leading to the unification of several partial-area nodes from several area nodes into one larger partial-area node. This can be observed, for example, by looking at the roots *Benign Muscle Neoplasm* (11), *Benign Brain Neoplasm* (15) and *Benign Female Reproductive System Neoplasm* (11).

Table 6
Example concepts with errors confirmed by NCIt curators

Concept	Confirmed Error	Correction
<i>Benign Buccal Mucosa Neoplasm</i>	Missing roles: <i>Disease Has Finding</i> with targets <i>Benign Cellular Infiltrate</i> and <i>Indolent Clinical Course</i>	Add these two roles
<i>Granulosa Cell Tumor</i>	Missing role: <i>Disease Has Normal Cell Origin</i> with the target <i>Granulosa Cell</i>	Add the role
<i>High Grade Vaginal Intraepithelial Neoplasia</i>	Missing role: <i>Disease Has Primary Anatomic Site</i> with the target <i>Vagina</i>	Add the role
<i>Human Papillomavirus-Related Malignant Neoplasm in AIDS Patient</i>	Missing role: <i>Disease Has Associated Disease</i> with the target <i>Acquired Immunodeficiency Syndrome</i>	Add the role
<i>Reproductive Endocrine Neoplasm</i>	Missing parent: <i>Endocrine Neoplasm</i>	Add IS-A to <i>Endocrine Neoplasm</i>

We submitted 76 consensus erroneous concepts to the NCIt editorial team. The NCIt team confirmed 17 erroneous concepts (22.4%), of which only one concept is from a large partial-area (with size 53) and the other 16 concepts are from small partial-areas. The NCIt team did not review any other concepts from the 190 concept sample and thus their review cannot be considered an alternative QA study. Table 6 lists five example concepts with errors that were confirmed and corrected by the NCIt team. Table 7 shows another four concepts that were reported as having errors by our domain experts, which were not corrected by the NCIt team. The third column in Table 7 reports the NCIt team's reasons for not correcting the concepts, while the fourth column presents our counter arguments, explaining why we nevertheless consider these legitimate errors. Table 8 shows the distribution of concepts into small, medium and large partial-areas.

5. Discussion

We have developed methodologies that are effective in uncovering ontology errors, caused unintentionally by curators and editors of ontologies as a result of making logical mistakes or of overlooking opportunities to provide more information. There are good reasons to employ an external team of professional domain experts for QA tasks, who might discover different logical errors or notice other opportunities for improvement than the creators of the ontologies.

One would imagine that such errors are so unpredictable that they are randomly distributed over the ontology, and therefore there is no way to find errors without also reviewing a large number of concepts without errors. This could be a very discouraging situation, in view of the scarce QA resources available for ontology maintenance. However, in this study we have demonstrated just the opposite. We have shown that although modeling errors are caused by unpredictable human behavior, nevertheless, there tend to be locations with higher than usual error concentrations in a specific ontology that can be identified algorithmically. QA review in these portions would result in a high QA yield, in terms of the ratio of the number of corrected concepts to the number of reviewed concepts.

For exposing such sets, the SABOC team utilizes Abstraction Networks such as the partial-area taxonomy, which can summarize an ontology by grouping together concepts with similar structure and/or semantics. An Abstraction Network offers an alternative compact visualization of an ontology's structure and content, which helps in detecting various anomalies not visible in the structure of the ontology itself. As we explained in the Methods section, one of the potential anomalies that manifests itself in the partial-area taxonomy, is the existence of small partial-areas.

Table 7

Example concepts with errors not corrected by NCIt curators

Concept	Reported Error	Curator's Reaction	Counter Argument
<i>High Grade Prostatic Intraepithelial Neoplasia</i>	Missing role: <i>Disease Has Finding</i> with the target <i>High Grade Lesion</i>	This concept already has the role <i>Disease Is Grade</i> with the target <i>High Grade</i> .	The role <i>Disease Is Grade</i> pointing to <i>High Grade</i> only hints at the existence of the role <i>Disease Has Finding</i> pointing to <i>High Grade Lesion</i> . In an ontology, information should be explicit so it is usable for computers. Besides, some concepts in NCIt have both these roles.
<i>Benign Skeletal Muscle Neoplasm</i>	Missing role: <i>Disease Has Associated Anatomic Site</i> with the target <i>Skeletal Muscle Tissue</i>	This concept already has the role <i>Disease Has Normal Tissue Origin</i> with the target <i>Skeletal Muscle Tissue</i> .	Similarly as above, the suggested role is not implied by the existing one.
<i>Lymphoplasmacyte-Rich Meningioma</i>	Missing role: <i>Disease Has Primary Anatomic Site</i> with the target <i>Meninges</i>	This concept already has the role <i>Disease Has Normal Tissue Origin</i> with the target <i>Meninges</i> .	Same as previous. However, some concepts in NCIt, (e.g., <i>Benign Meningioma</i>), have both the two roles <i>Disease Has Primary Anatomic Site</i> and <i>Disease Has Normal Tissue Origin</i> pointing to <i>Meninges</i> .
<i>Clear Cell Squamous Cell Skin Carcinoma</i>	Missing parent: <i>Primary Malignant Neoplasm</i>	<i>Primary Malignant Neoplasm</i> is under "Neoplasm by Special Category." The NCIt team had decided not to populate such terms with specific histopathologies and not to model/define such terms.	The NCIt curator agrees with the correction and does not implement it due to NCIt editorial policy.

Table 8

The neoplasm concept distribution according to partial-area size

Partial-area size	# of Partial-areas	Total # of Concepts	% of Concepts
1–10	4,762	6,581	80.59%
11–19	44	642	7.86%
≥20	18	1,014	12.42%
Total:	4,824	8,166	

A small partial-area represents a unique grouping of just a few concepts with a very specific structure and semantics out of thousands of concepts of an ontology. Sometimes this unique grouping correctly reflects the reality of the application domain, but often it is the result of a modeling mistake. In the current research, we have examined this phenomenon in the *Neoplasm* subhierarchy of NCIt, the most important part of NCIt for the mission of NCI.

When we previously discovered the phenomenon of a high likelihood of errors for small partial-areas (Halper et al., 2007; Min et al., 2006), we established that there is a higher percentage of errors in small partial-areas compared to larger partial-areas, but without showing statistical significance. However, in other studies we have found that this methodology is not universally applicable. For example, Min et al. (2008) found that a very large majority of the partial-area taxonomy for the NCIt Gene hierarchy

consisted of partial-areas of one concept – the genes themselves were leaves of this hierarchy and some roles were first introduced only at the gene concepts. Hence, for the modeling of the *Gene* hierarchy of NCIt, small partial-areas were *not* uncommon, but a common way of modeling concepts. Therefore, a small partial-area was not an indicator of a high probability of errors for the *Gene* hierarchy. In other words, what is “uncommon” depends on the context provided by a specific hierarchy.

Similarly, in our QA study of the *Biological Process* hierarchy of the Gene Ontology (GO) (Ochs et al., 2016), there was a large number of small partial-areas in the GO partial-area taxonomy. Again, for the modeling of GO, many roles were introduced at the leaf concepts, creating small partial-areas of a single concept, making them the norm rather than an anomaly.

We observe that both of our previous studies regarding small partial-areas were on small hierarchies with a medium number of roles. One was the *Biological Process* hierarchy of NCIt, which had 589 concepts and seven roles in 2004 when the research was conducted (Min et al., 2006). The second was the *Specimen* hierarchy of SNOMED CT with 1,056 concepts and five relationships in the 2004 release of SNOMED CT (Halper et al., 2007). In view of the ontologies for which small partial-areas did *not* indicate more errors than larger partial-areas, we decided to conduct the current study on the *Neoplasm* subhierarchy of 8,166 concepts in the *Disease, Disorder or Finding* hierarchy of 25,360 concepts. This hierarchy has 29 different roles.

One should realize that a higher number of roles defined for a hierarchy may increase the number of partial-areas and hence decreases their sizes, since each extra role has the potential to further divide the existing partial-areas in the partial-area taxonomy obtained for an ontology with one fewer role. Furthermore, the current research was only enabled by recently introducing the notion of a partial-area taxonomy limited to a subhierarchy of an ontology hierarchy (Ochs et al., 2015). This new notion enabled us to limit the study to the *Neoplasm* subhierarchy. Previously it was impossible to derive a partial-area taxonomy for a subhierarchy.

A natural question arises out of our results. What percentage of the Neoplasm concepts is located in small partial-areas for which our study found a 44% error rate? Table 8 shows the distribution of concepts into small, medium and large partial-areas. More than 80% of the concepts are in small partial-areas (of those 3,997 (48.94%) are in partial-areas of one concept). In contrast, there are only 12.41% in the large partial-areas exhibiting a 22% error rate. Hence, reviewing the small partial-areas will lead to the discovery of a majority of the errors, with a high yield.

We note that the sum of the numbers of concepts of the different size categories add up to more than the 8,166 Neoplasm concepts. The reason is that partial-areas are not necessarily disjoint and therefore some concepts appear in more than one partial-area and are counted more than once. This phenomenon was explored by Wang et al. (2012).

We stress that one must distinguish the QA analysis by an external domain expert from the secondary review of these results by the ontology’s curators, performed for the purpose of correcting the ontology. The external domain expert is free from any commitment to any *a priori* modeling decisions made during the initial design and the life cycle of the ontology. Furthermore, the NCIt team does not have written formal guidelines describing the rules they follow in their modeling. (The external reviewers requested such guidelines and were informed about their absence.) The duty of an external expert is to question any modeling details that appear erroneous or inconsistent according to generally accepted ontological principles. A secondary function of an external expert is to suggest corrections.

As indicated by Column 4 of Table 7, the NCIt curator did not object to the findings of the external reviewers, but due to various editorial rules, decided not to implement these suggested corrections. The external reviewers asked the curator to supply them with the editorial rules underlying the design of

the hierarchy, but these were not available in writing. Thus, the way to measure the effectiveness of the techniques suggested in this paper should be based on the statistics of the findings of the external reviewers. It should not be based on the degree to which the NCIt team accepts these findings.

The findings of this study have to be contrasted with prior work. We performed QA of several small OWL ontologies, all hosted by the BioPortal system (Whetzel et al., 2011) of the National Center for Biomedical Ontology (NCBO) (Musen et al., 2012). These ontologies included the Ontology of Clinical Research (OCRe), the Sleep Domain Ontology (SDO), the Drug Discovery Investigations Ontology (DDI), and the Cancer Chemoprevention Ontology (CanCo). The sizes of these ontologies are measured in the hundreds of concepts. For all of them, small partial-areas were the norm rather than the exception, and small-partial areas did not exhibit higher error rates than larger ones. This stands in marked contrast to the present study on a large ontology hierarchy. One goal of future work is to find formal ways to characterize the ontologies for which small partial-areas can uncover more errors. The size of the ontology appears to be a factor, but there may well be other hidden variables. Characterizing the properties of ontologies and of hierarchies within ontologies, for which small partial-areas exhibit higher error concentrations would enable curators to better manage their limited QA resources.

Another idea for future research is based on the results of the group auditing and error correction propagation techniques described in Sections 3.2 and 3.3. We will investigate the efficacy of auditing only the roots of the small partial-areas. For all the erroneous roots found, the results will be used for group auditing and corrections will be propagated to down the partial-area taxonomy. Working in this way is expected to increase the QA yield obtained, compared to the yield for any concepts from small partial-areas. On the other hand, this technique will not expose errors appearing in the middle of a partial-area. Future research will explore the trade-off between the advantages and disadvantages of “root-based auditing.”

In this paper, we assumed that medium-sized partial-areas have between 11 and 19 concepts. We will experiment with different thresholds between small and medium-sized partial-areas and between medium-sized and large partial-areas to optimize the definition of small partial-areas with high error rates, and large partial-areas with low error rates.

6. Conclusions

Errors in the NCIt *Neoplasm* subhierarchy are not uniformly distributed. Constructing a specific Abstraction Network, called a partial-area taxonomy, of the *Neoplasm* subhierarchy (of the *Disease, Disorder or Finding* hierarchy), results in sets of concepts at two granularities, namely as areas and as partial-areas. Applying our meta-methodology of looking for uncommon modeling of concepts, which is reflected in small sets of concepts in the partial-area taxonomy, we found a significantly larger percentage of erroneous concepts than in a control group of concepts from large partial-areas. The error rate for small partial-areas (44.7%) was twice as large as the error rate for large partial-areas (22.5%).

Furthermore, we demonstrated that group-based auditing, using groups constituted in the partial-area taxonomy, supports easy discovery of additional erroneous concepts at the same level of the partial-area taxonomy (in fact, in the same partial-area). By error correction propagation, additional errors at lower levels in the partial-area taxonomy were also found and corrected with minimal additional effort. On a more general level, we conclude that Abstraction Networks were again successful in aiding the process of discovering “suspicious” concepts. Three human auditors, in a two-step process, suggested corrections for erroneous concepts. The most common error type in our sample was “missing role,” with “missing parent” as a distant second.

Acknowledgements

We thank Sherri De Coronado, the manager of the NCIt team, and Nicholas Sioutos, a curator of the NCIt team, for their feedback and remarks regarding the results of this study.

Research reported in this publication was partially supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA190779. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Arabandi, S., Ogbuji, C., Redline, S., Chervin, R., Boero, J., Benca, R. & Zhang, G. (2010). Developing a Sleep Domain Ontology. *AMIA Clinical Research Informatics Summit*.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al. (2000). Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nat Genet*, 25(1), 25–29. doi:10.1038/75556.
- Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res*, 32(Database issue), D267–D270. doi:10.1093/nar/gkh061.
- Bodenreider, O. (2008). Biomedical ontologies in action: Role in knowledge management, data integration and decision support. *Yearb Med Inform*, 67–79.
- Campbell, J.R., Campbell, W.S., Hickman, H., Pedersen, J. & McClay, J. (2015). Employing complex polyhierarchical ontologies and promoting interoperability of i2b2 data systems. *AMIA Annu Symp Proc*, 2015, 359–365.
- Ceusters, W., Smith, B. & Goldberg, L. (2005). A terminological and ontological analysis of the NCI thesaurus. *Methods Inf Med*, 44(4), 498–507.
- Dalkey, N. & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458–467. doi:10.1287/mnsc.9.3.458.
- De Coronado, S., Haber, M.W., Sioutos, N., Tuttle, M.S. & Wright, L.W. (2004). NCI thesaurus: Using science-based terminology to integrate cancer research results. *Stud Health Technol Inform*, 107(Pt 1), 33–37.
- De Coronado, S., Wright, L.W., Fragoso, G., Haber, M.W., Hahn-Dantona, E.A., Hartel, F.W., et al. (2009). The NCI thesaurus quality assurance life cycle. *J Biomed Inform*, 42(3), 530–539. doi:10.1016/j.jbi.2009.01.003.
- Donfack, G.V., Bertaud, G.V., Duvauferrier, R., Bourde, A., Morelli, J. & Lasbleiz, J. (2012). Ontology driven decision support systems for medical diagnosis – an interactive form for consultation in patients with plasma cell disease. *Stud Health Technol Inform*, 180, 108–112.
- EVS Use and Collaborations (NCIt Internal Document) (2014). August 19. Retrieved March 27, 2017 from, [https://wiki.nci.nih.gov/display/EVS/EVS+Use+and+Collaborations+\(Entire+Document\)](https://wiki.nci.nih.gov/display/EVS/EVS+Use+and+Collaborations+(Entire+Document)).
- Fragoso, G., De Coronado, S., Haber, M., Hartel, F. & Wright, L. (2004). Overview and utilization of the NCI thesaurus. *Comp Funct Genomics*, 5(8), 648–654. doi:10.1002/cfg.445.
- Geller, J., He, Z., Perl, Y., Morrey, C.P. & Xu, J. (2013). Rule-based support system for multiple UMLS semantic type assignments. *J Biomed Inform*, 46(1), 97–110. doi:10.1016/j.jbi.2012.09.007.
- Geller, J., Perl, Y., Halper, M. & Cornet, R. (2009). Special issue on auditing of terminologies. *J Biomed Inform*, 42(3), 407–411. doi:10.1016/j.jbi.2009.04.006.
- Good, P.I. (2005). *Permutation, Parametric, and Bootstrap Tests of Hypotheses: A Practical Guide to Resampling* (3rd ed.). New York, NY: Springer.
- Gu, H., Elhanan, G., Perl, Y., Hripcsak, G., Cimino, J.J., Xu, J., et al. (2012). A study of terminology auditors' performance for UMLS semantic type assignments. *J Biomed Inform*, 45(6), 1042–1048. doi:10.1016/j.jbi.2012.05.006.
- Gu, H., Hripcsak, G., Chen, Y., Morrey, C.P., Elhanan, G., Cimino, J.J., et al. (2007). Evaluation of a UMLS auditing process of semantic type assignments. In *AMIA Annu Symp Proc* (pp. 294–298).
- Gu, H., Perl, Y., Elhanan, G., Min, H., Zhang, L. & Peng, Y. (2004). Auditing concept categorizations in the UMLS. *Artif Intell Med*, 31(1), 29–44. doi:10.1016/j.artmed.2004.02.002.
- Gu, H., Perl, Y., Geller, J., Halper, M., Liu, L.M. & Cimino, J. (2000). Representing the UMLS as an object-oriented database: Modeling issues and advantages. *J Am Med Inform Assoc*, 7(1), 66–80. doi:10.1136/jamia.2000.0070066.
- Halper, M., Gu, H., Perl, Y. & Ochs, C. (2015). Abstraction networks for terminologies: Supporting management of “big knowledge”. *Artif Intell Med*, 64(1), 1–16. doi:10.1016/j.artmed.2015.03.005.
- Halper, M., Morrey, C.P., Chen, Y., Elhanan, G., Hripcsak, G. & Perl, Y. (2011). Auditing hierarchical cycles to locate other inconsistencies in the UMLS. *AMIA Annu Symp Proc*, 2011, 529–536.

- Halper, M., Wang, Y., Min, H., Chen, Y., Hripcsak, G., Perl, Y. & Spackman, K.A. (2007). Analysis of error concentrations in SNOMED. *AMIA Annu Symp Proc*, 314–318.
- He, Z., Morrey, C.P., Perl, Y., Elhanan, G., Chen, L., Chen, Y. & Geller, J. (2014). Sculpting the UMLS refined semantic network. *Online J Public Health Inform*, 6(2), e181. doi:[10.5210/objphi.v6i2.5412](https://doi.org/10.5210/objphi.v6i2.5412).
- He, Z., Ochs, C., Agrawal, A., Perl, Y., Zeginis, D., Tarabanis, K., et al. (2013a). A family-based framework for supporting quality assurance of biomedical ontologies in BioPortal. In *AMIA Annu Symp Proc*, 2013 (pp. 581–590).
- He, Z., Ochs, C., Soldatova, L., Perl, Y., Arabandi, S. & Geller, J. (2013b). Auditing redundant import in reuse of a top level ontology for the drug discovery investigations ontology. In *Proceedings of the 4th International Conference on Biomedical Ontology*.
- Hochheiser, H., Castine, M., Harris, D., Savova, G. & Jacobson, R.S. (2016). An information model for computable cancer phenotypes. *BMC Med Inform Decis Mak*, 16(1), 121. doi:[10.1186/s12911-016-0358-4](https://doi.org/10.1186/s12911-016-0358-4).
- Humphreys, B.L., Lindberg, D.A., Schoolman, H.M. & Barnett, G.O. (1998). The Unified Medical Language System: An informatics research collaboration. *J Am Med Inform Assoc*, 5(1), 1–11. doi:[10.1136/jamia.1998.0050001](https://doi.org/10.1136/jamia.1998.0050001).
- Jiang, G., Sohn, S., Zimmermann, M.T., Wang, C., Liu, H. & Chute, C.G. (2015). Drug normalization for cancer therapeutic and druggable genome target discovery. *AMIA Jt Summits Transl Sci Proc*, 2015, 72–76.
- McCray, A.T. (2003). An upper-level ontology for the biomedical domain. *Comp Funct Genomics*, 4(1), 80–84. doi:[10.1002/cfg.255](https://doi.org/10.1002/cfg.255).
- McCray, A.T. & Nelson, S.J. (1995). The representation of meaning in the UMLS. *Methods Inf Med*, 34(1–2), 193–201.
- Min, H., Cohen, B., Halper, M., Oren, M. & Perl, Y. (2008). Detecting role errors in the gene hierarchy of the NCI thesaurus. *Cancer Inform*, 6, 293–313.
- Min, H., Perl, Y., Chen, Y., Halper, M., Geller, J. & Wang, Y. (2006). Auditing as part of the terminology design life cycle. *J Am Med Inform Assoc*, 13(6), 676–690. doi:[10.1197/jamia.M2036](https://doi.org/10.1197/jamia.M2036).
- Morrey, C.P., Geller, J., Halper, M. & Perl, Y. (2009). The Neighborhood Auditing Tool: A hybrid interface for auditing the UMLS. *J Biomed Inform*, 42(3), 468–489. doi:[10.1016/j.jbi.2009.01.006](https://doi.org/10.1016/j.jbi.2009.01.006).
- Morrey, C.P., Perl, Y., Halper, M., Chen, L. & Gu, H.H. (2012). A chemical specialty semantic network for the Unified Medical Language System. *J Cheminform*, 4(1), 9. doi:[10.1186/1758-2946-4-9](https://doi.org/10.1186/1758-2946-4-9).
- Mougin, F. & Bodenreider, O. (2008). Auditing the NCI thesaurus with semantic web technologies. In *AMIA Annu Symp Proc* (pp. 500–504).
- Musen, M.A., Noy, N.F., Shah, N.H., Whetzel, P.L., Chute, C.G., Story, M.A., et al. (2012). The National Center for Biomedical Ontology. *J Am Med Inform Assoc*, 19(2), 190–195. doi:[10.1136/amiainl-2011-000523](https://doi.org/10.1136/amiainl-2011-000523).
- Ochs, C., Agrawal, A., Perl, Y., Halper, M., Tu, S.W., Carini, S., et al. (2012). Deriving an abstraction network to support quality assurance in OCRE. In *AMIA Annu Symp Proc*, 2012 (pp. 681–689).
- Ochs, C., Geller, J., Perl, Y., Chen, Y., Xu, J., Min, H., et al. (2015). Scalable quality assurance for large SNOMED CT hierarchies using subject-based subtaxonomies. *J Am Med Inform Assoc*, 22(3), 507–518. doi:[10.1136/amiainl-2014-003151](https://doi.org/10.1136/amiainl-2014-003151).
- Ochs, C., He, Z., Perl, Y., Arabandi, S., Halper, M. & Geller, J. (2013). Choosing the granularity of abstraction networks for orientation and quality assurance of the sleep domain ontology. In *Proceedings of the 4th International Conference on Biomedical Ontology*.
- Ochs, C., Perl, Y., Halper, M., Geller, J. & Lomax, J. (2016). Quality assurance of the gene ontology using abstraction networks. *J Bioinform Comput Biol*, 14(3), 1642001. doi:[10.1142/S0219720016420014](https://doi.org/10.1142/S0219720016420014).
- Park, Y.R., Kim, J.J., Yoon, Y.J., Yoon, Y.K., Koo, H.Y., Hong, Y.M., et al. (2016). Establishment of Kawasaki disease database based on metadata standard. *Database (Oxford)*, 2016. doi:[10.1093/database/baw109](https://doi.org/10.1093/database/baw109).
- Qi, D., King, R.D., Hopkins, A.L., Bickerton, G.R. & Soldatova, L.N. (2010). An ontology for description of drug discovery investigations. *J Integr Bioinform*, 7(3). doi:[10.2390/biecoll-jib-2010-126](https://doi.org/10.2390/biecoll-jib-2010-126).
- Reed, T.L. & Kaufman-Rivi, D. (2010). FDA adverse Event Problem Codes: Standardizing the classification of device and patient problems associated with medical device use. *Biomed Instrum Technol*, 44(3), 248–256. doi:[10.2345/0899-8205-44.3.248](https://doi.org/10.2345/0899-8205-44.3.248).
- Rubin, D.L., Shah, N.H. & Noy, N.F. (2008). Biomedical ontologies: A functional perspective. *Brief Bioinform*, 9(1), 75–90. doi:[10.1093/bib/bbm059](https://doi.org/10.1093/bib/bbm059).
- Sim, I., Tu, S.W., Carini, S., Lehmann, H.P., Pollock, B.H., Peleg, M. & Wittkowski, K.M. (2014). The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research. *J Biomed Inform*, 52, 78–91. doi:[10.1016/j.jbi.2013.11.002](https://doi.org/10.1016/j.jbi.2013.11.002).
- Stearns, M.Q., Price, C., Spackman, K.A. & Wang, A.Y. (2001). SNOMED clinical terms: Overview of the development process and project status. In *Proc AMIA Symp* (pp. 662–666).
- Wang, Y., Halper, M., Min, H., Perl, Y., Chen, Y. & Spackman, K.A. (2007). Structural methodologies for auditing SNOMED. *J Biomed Inform*, 40(5), 561–581. doi:[10.1016/j.jbi.2006.12.003](https://doi.org/10.1016/j.jbi.2006.12.003).
- Wang, Y., Halper, M., Wei, D., Gu, H., Perl, Y., Xu, J., et al. (2012). Auditing complex concepts of SNOMED using a refined hierarchical abstraction network. *J Biomed Inform*, 45(1), 1–14. doi:[10.1016/j.jbi.2011.08.016](https://doi.org/10.1016/j.jbi.2011.08.016).

- Wei, D. & Bodenreider, O. (2010). Using the abstraction network in complement to description logics for quality assurance in biomedical terminologies – a case study in SNOMED CT. *Stud Health Technol Inform*, 160(Pt 2), 1070–1074.
- Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T. & Musen, M.A. (2011). BioPortal: Enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res*, 39 (Web Server issue), W541–545. doi:[10.1093/nar/gkr469](https://doi.org/10.1093/nar/gkr469).
- Zeginis, D., Hasnain, A., Loutas, N., Deus, H.F., Fox, R. & Tarabanis, K. (2014). A collaborative methodology for developing a semantic model for interlinking cancer chemoprevention linked-data sources. *Semantic Web*, 5(2), 127–142.
- Zhu, X., Fan, J.W., Baorto, D.M., Weng, C. & Cimino, J. (2009). A review of auditing methods applied to the content of controlled biomedical terminologies. *J Biomed Inform*, 42(3), 413–425. doi:[10.1016/j.jbi.2009.03.003](https://doi.org/10.1016/j.jbi.2009.03.003).