

Stop-word based contextual auditing to identify inconsistencies in SNOMED

Rashmi Burse, Gavin McArdle, and Michela Bertolotto

University College Dublin, Belfield, Dublin 4, Ireland
rashmi.burse@ucdconnect.ie

Abstract. SNOMED is one of the most widely adopted Clinical Terminology systems. However, incomplete representations and modelling inconsistencies in SNOMED are preventing healthcare applications from exploiting its full potential. This paper presents a novel stop-word based contextual auditing method to identify potential inconsistencies in the modelling of SNOMED concepts. The results of a pilot study method show promising potential with this method. The percentage of identified missing attribute relationships using this method is as high as 69.56% and for identified missing hierarchical relationships it is 28.26%. The auditing method proposed in this paper can act as a supplementary Quality Assurance check in the International Health Terminology Standards Development Organization's effort to improve the quality of SNOMED.

Keywords: SNOMED · Quality Assurance · Lexical Auditing

1 Introduction

Incomplete, inconsistent and erroneous representations of Clinical Terminology (CT) systems limit their expressiveness and have a variety of repercussions including retrieval of incomplete or incorrect result sets. Missing relationships result in the existence of partially defined concepts which obstruct the divulgence of rich inferential knowledge. For example, in the International Edition of March 2020 SNOMED version, the concept *Insomnia with sleep apnea (disorder)* has only one parent, *Insomnia (disorder)*. The hierarchical link to *Sleep apnea (disorder)* is absent. *Sleep apnea (disorder)* has a role group containing three attribute relationships which are missing from the concept *Insomnia with sleep apnea (disorder)*, thus preventing it to capture all relevant information to define this condition. If someone executes a query to retrieve all patients suffering from *sleep apnea (disorder)*, the patients suffering from *Insomnia with sleep apnea (disorder)* would not be retrieved due to the missing hierarchical relationship between *sleep apnea (disorder)* and *Insomnia with sleep apnea (disorder)*. This will yield inaccurate partial results. Given the critical nature of medical data, effective Quality Assurance (QA) of CT systems is imperative.¹

¹ Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

However, the development of effective auditing methods for the QA of CT systems is a major challenge and an ongoing process in the health-informatics domain. In spite of continual research efforts, the healthcare community is still striving to hone its auditing techniques for two major reasons: (a) the huge size of CT systems makes it impractical to audit each and every concept manually. (b) the diverse nature of clinical data has led to a variety of conflicting modelling styles making it impossible to develop a "one size fits all" solution that can be applied to all CT systems. Taking into consideration these constraints, the best way forward is to develop efficient auditing techniques that highlight concentrated erroneous regions in a CT system. Such areas can then be presented to authors and curators of a CT system for manual inspection. The main objective of such techniques is to direct the limited available resources to highly concentrated erroneous areas and identify maximum number of inconsistencies with minimal effort.

With this objective, we present a novel method based on lexical analysis of concept names containing stop-words. It is our hypothesis that stop-words which have been disregarded by other lexical auditing methods can prove to be rich sources of information to identify problematic areas. The pilot version of this method is restricted to the stop-words "and" and "with" due to their conjunctive nature. However, we plan to expand our analysis to other stop-words in the future. The proposed method identifies two types of inconsistencies: missing hierarchical relationships (i.e., if a SNOMED concept exists, which is lexically equivalent or a lexical variant of any of the subjects appearing before or after the stop-word and is not assigned as a parent of the concept) and missing attribute relationships (i.e., in the case of a missing hierarchical relationship, if the attribute relationship(s) of the identified lexically ideal parent is/are not included as a role group in the modeling of the concept). The proposed method promotes semantic completeness by identifying missing attribute relationships to refine a concept and ensures consistency in structural modelling by identifying missing hierarchical relationships. An additional advantage of our method over other auditing methods is that it not only identifies inconsistencies but also provides a potential list of suggestive corrections for each identified inconsistency. The aim of our method is to highlight areas with a high concentration of errors in order to save time and effort of experts and curators on manual auditing.

2 Related Work

Bodenreider et al. [15] developed a method to identify missing elements in SNOMED by targeting concepts containing binary antonymous adjectives such as (acute, chronic), (unilateral, bilateral), (primary, secondary), and (acquired, congenital). The proposed method extracted adjectival modifiers from the targeted concepts ([MOD][CONTEXT]) and created new terms by experimenting with various combinations of modifiers and contexts. Bodenreider et al. [14] exploited the lexical features of concepts to identify missing hyponymic relationships. The method selected concepts conforming to a modifier+noun form

([MOD][NOUN]), where modifier was usually an adjectival modifier further describing the noun. They intuitively assumed that modifier+noun should be a hyponym of the noun, e.g. *acute appendicitis* should be a child of *appendicitis*, and identified missing hyponymic relationships. Pacheco et al. [20] assumed that non-attributed concepts were underspecified and employed a semantic indexing method to suggest attribute relationships for such concepts. The method derived sub-words from a non-attributed concept's Fully Specified Name (FSN) with the help of MorphoSaurus [18]. The derived sub-words were compared with the concept's parent(s). Common sub-words appearing both in child and parent concept were eliminated. The concepts containing the remaining sub-words were then searched and chosen as eligible candidates to refine the non-attributed concept.

Agrawal and Elhanan [5] examined five types of inconsistencies among concepts whose FSNs were lexically similar, i.e., differed by only one word. The method created similarity sets consisting of concepts that differed from a base descriptor by one word. E.g. for the base descriptor "upper limb stretching", *Prophylactic upper limb stretching (procedure)*, *Therapeutic upper limb stretching (procedure)*, and *Prophylactic lower limb stretching (procedure)* constituted a similarity set. The method was applied to Procedure sub-hierarchy of SNOMED. 5 samples each consisting of 50 similarity sets were created and each sample was examined for hierarchical, attribute assignment, attribute target value, group, and definitional inconsistencies. Bodenreider [13] claimed that the root cause for all inconsistencies in CT systems was concepts modeled with faulty logical definitions. With this notion they recreated logical definitions from the lexical features of a concept name and inferred hierarchical relationships among these newly defined concepts. The newly obtained hierarchy was then compared with the original SNOMED hierarchy to detect differences. Schulz et al. [22] detected ambiguities in hierarchy tags, attribute relationships, and IS-A relationships based on the lexical features of SNOMED concepts and made some valuable suggestions for the curators of SNOMED. Rector and Iannone [21] focused on finding concepts from the findings and diseases sub-hierarchies of SNOMED that should be classified as chronic or acute according to CORE problem list but currently are not and studied the effect of this misclassification on post-coordination queries. Ceusters et al. [16] scrutinized concepts containing negation words like absence, negation, and not and misclassification caused due to these words. They introduced four categories into which negative relationships can be classified, suggested that SNOMED should be aligned with an Upper Level Ontology (ULO) like Basic Formal Ontology (BFO), and introduced a new "lacks" relationship to correctly classify such negative concepts.

Agrawal et al. [7] reported the results of a study that statistically concluded that the complexity and thereby the chances of identifying errors increases with the length (number of words) of a concept name and the number of parents of a concept. Agrawal [4] proposed an auditing method based on the hypothesis that if two concepts are lexically similar then their structural and logical modeling

should also be similar. E.g. the concepts *Acute injury of anterior cruciate ligament (disorder)* and *Acute injury of posterior cruciate ligament (disorder)* are lexically similar as they differ by only one word and hence have similar structural and logical modelling. Both concepts have the same number of hierarchical relationships, same number and type of attribute relationships differing only in the target values (anterior and posterior). Many variations of this method, including simple similarity sets [6, 12], positional similarity sets [8, 9], and employing machine learning tools to create similarity sets [10, 11] were developed and applied to different versions and sub-hierarchies of SNOMED. Cui et al. [17] proposed a hybrid method combining the structural and lexical aspects of a CT system and identified four lexical patterns in non-lattice subgraphs that suggested potential missing hierarchical relationships and potential missing concepts.

To summarize, all the lexical auditing methods applied so far work on one of the following principles (a) counting the length of a concept name to estimate its complexity and thereby calculate the probability of potential inconsistencies harbored by it; (b) performing lexico-syntactic and morphosyntactic analysis on the concept names to identify missing concepts/relationships; (c) applying normalization techniques and LVG algorithms to deal with variation in concept names; (d) looking for lexical similarity among concept names to check for inconsistencies in their structural and logical modelling.

The intent and focus of all the aforementioned methods is on medical jargons and their lexical variants. As a result, these methods scrutinized fixed parts of speech like adjectival modifiers, nouns, and verbs and found repeatedly occurring stop-words like “and”, “or”, “with” etc. to be a hindrance. To improve the performance efficiency of their algorithms, these methods ignored a list of such stop-words [3]. These stop-words that are disregarded and eliminated by all the aforementioned studies can prove to be rich sources of information to identify problematic areas. They can serve as effective indicators to identify concepts harboring potential inconsistencies. The stop-word list [3] eliminated by these studies serves as a major motivation for our approach. In this work we present a unique method that targets concepts containing stop-words, “and” and “with”, to identify two types of inconsistencies: missing hierarchical relationships and missing attribute relationships. The pilot version of this method is restricted to the stop-words “and” and “with” due to their conjunctive nature. However, we plan to expand our analysis to other stop-words [3] in the future. To the best of our knowledge, there is no lexical method developed so far that targets stop-words to audit CT systems.

3 Materials and Method

3.1 Materials

In this pilot study, the proposed method will be applied to the *Disorder* sub-hierarchy of SNOMED’s March 2020 International Edition. However, the proposed method is quite generic and can be applied to other hierarchies of

SNOMED as well as other CT systems. We have chosen this sub-hierarchy because after performing a preliminary inspection, we found many concepts in the *disorder* sub-hierarchy containing the stop-words “and” and “with” that were either missing hierarchical relationships or were assigned inconsistent hierarchical relationships that varied in granularity and were missing attribute relationships. There are almost 7000 eligible concepts, containing “and” or “with”, that need to be systematically assessed and it is our hypothesis that the proposed method will highlight erroneous concepts that require manual auditing.

3.2 Method

The proposed method is based on four assumptions and identifies two types of inconsistencies. Lexical variants in this work are considered to be concept FSNs conforming to the lexical structure “subject + syndrome” and terms appearing before and after “and” or “with” will hereafter be referred to as subjects. Inconsistencies are defined as follows:

Missing hierarchical relationship: If a SNOMED concept exists, which is lexically equivalent or a lexical variant of any of the subjects and is not assigned as a parent of the concept.

Missing attribute relationship (role group): In case of a missing hierarchical relationship, if the attribute relationship(s) of the identified lexically ideal parent is/are not included as a role group in the modeling of the concept.

The assumptions made in this study are based on the observation that concepts containing “and” and “with” are expected to have at least two parents and at least two role groups. The first assumption is also supported by a semantic rule proposed during the early formative years of SNOMED [19]. Mendonca et al. [19] conducted a thorough analysis of SNOMED concepts containing conjunctions like “and”, “and/or”, “either/or”, “neither/nor” and came to the conclusion that if a SNOMED concept contains the word “and”, it should be treated as a “logical and” and the properties of the subjects appearing before and after the conjunction must be present in the concept. All other cases that entertain the idea of exclusivity allowing the presence of either one or both subjects should be represented using the more lenient “and/or” conjunction.

Fig. 1 illustrates the example of a concept *Pneumonia and influenza (disorder)* which has two parents *influenza(disorder)* and *Pneumonia(disorder)*. The names of the parents are lexically equivalent to the subjects. It has two role groups one belonging to each of the parent disorders, i.e. role group 1 containing three attribute relationships: *pathological process – infectious process*, *causative agent – influenza virus*, *finding site – structure of respiratory system* belonging to *influenza (disorder)* and role group 2 containing two attribute relationships: *associated morphology – Inflammation and consolidation*, *finding site – lung structure* belonging to *pneumonia (disorder)*. Fig. 2 illustrates the individual disorder concepts *pneumonia (disorder)* and *influenza (disorder)* along with their role groups. The diagrammatic representations of concepts are downloaded from IHTSDO’s SNOMED browser [2]. Based on this observation and the semantic rules mentioned in [19], we present Assumptions 1 and 2.

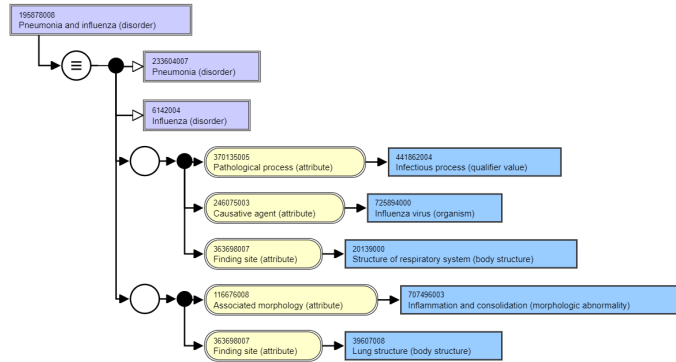


Fig. 1. Diagrammatic representation of SNOMED concept “*Pneumonia and influenza (disorder)*” SCTID: 195878008”

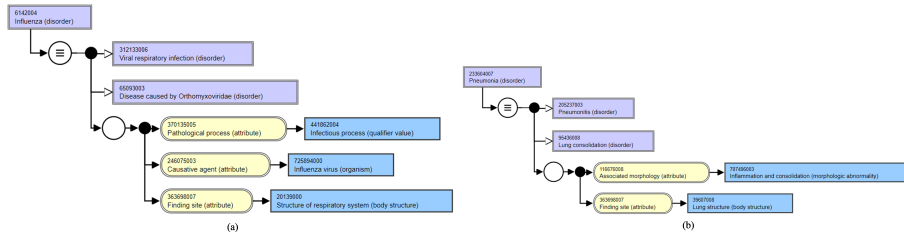


Fig. 2. (a) Diagrammatic representation of SNOMED concept “*Influenza (disorder)*” SCTID: 6142004” (b) Diagrammatic representation of SNOMED concept “*Pneumonia (disorder)*” SCTID:233604007”

Assumption 1 Concepts containing the stop-word “and” should have at least two parents and the parents must either be lexically equivalent or must be lexical variants of the subjects appearing before and after “and”.

Assumption 2 Concepts containing the stop-word “and” should have at least two role groups, and the role groups should be equivalent to the role groups of each individual concept corresponding to the subjects appearing before and after “and”.

Fig. 3 illustrates the example of a concept *Ornithosis with pneumonia (disorder)* which has four parents including *Ornithosis (disorder)* and *Pneumonia (disorder)* and two role groups, one for each individual disorder parent corresponding to the subject. Fig. 4 illustrates the individual concept *Ornithosis (disorder)* along with its role group. The other parent *Pneumonia (disorder)* along its role group is already illustrated in Fig. 2 (b). Based on this observation, we present Assumptions 3 and 4.

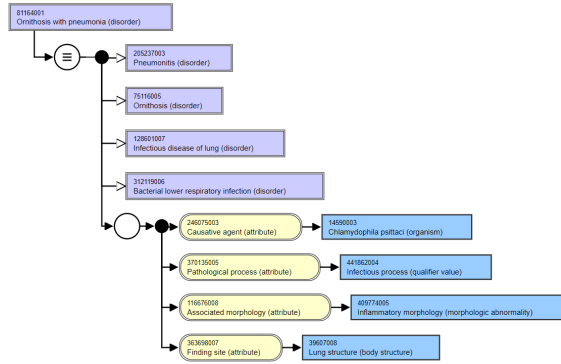


Fig. 3. Diagrammatic representation of SNOMED concept “*Ornithosis with pneumonia (disorder) SCTID:81164001*”

Assumption 3 Concepts containing the stop-word “with” should have at least two parents and the parents must either be lexically equivalent or must be lexical variants of the subjects appearing before and after “with”.

Assumption 4 Concepts containing the stop-word “with” should have at least two role groups, and the role groups should be equivalent to the role groups of each individual concept corresponding to the subjects appearing before and after “with”.

We formulated a set of rules based on the aforementioned assumptions which form the backbone of our algorithm. The developed algorithm identifies missing hierarchical relationships, missing attribute relationships, and also makes corrective suggestions by listing lexically ideal concepts using the four assumptions.

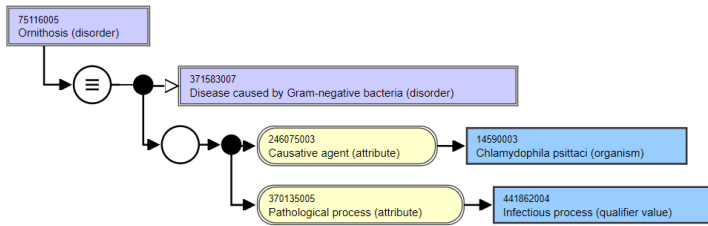


Fig. 4. Diagrammatic representation of SNOMED concept “*Ornithosis (disorder) SCTID: 75116005*”

4 Results and Discussion

Table 1 displays the number of eligible concepts containing the keywords “and” and “with” which were found in the *disorder* sub-hierarchy of SNOMED’s

Inter-national Edition March 2020 release. The pilot study is limited to concepts containing a maximum of three words (excluding the hierarchy tag, (disorder)) in their Fully Specified Names (FSNs). From Table 1, we can see that out of 6989 concepts containing stop-words “and” or “with”, 92 concepts have a maximum of three words in their FSN.

Table 1. Number of eligible concepts

Description	#
Total concepts in <i>Disorder</i> sub-hierarchy (active only)	76747
Concepts containing stop-words “and” and “with” (FSN length - any)	6989
Concepts containing stop-words “and” and “with” (FSN length - 3)	92

Out of the 92 concepts, 26 concepts (28.26%) were identified to be missing one or more parent(s) according to the lexical rules stated in assumptions 1- 4. Out of the 26 concepts, 3 concepts had all suggested parents that belonged to *finding* sub-hierarchy. Currently, these concepts are dropped from the analysis due to lack of medical expertise to check conformance with the guidelines [1], but will be covered in future work after developing appropriate rules for such cases.

Out of the 23 concepts, 16 concepts (69.56%) were found to be missing attribute relationships. Table 2 reports the statistics of the results related to missing hierarchical relationships and Table 3 reports the statistics of the results related to missing attribute relationships that were obtained by our method. In tables 2 and 3, the second column (#) displays the number of concepts belonging to the category described by the first column (Description), the third column (Percentage) displays the count in terms of percentage and the fourth and fifth columns display the “and” and “with” concept distribution of the count respectively. Table 4 lists the top three missing parents and missing attribute relationships identified by our method. In table 4, the first column represents the identified concept containing the stop-word “and” or “with”, second column displays the suggested missing hierarchical relationship, i.e. missing parent, and the third column represents its corresponding attribute relationship that should be ideally present but is missing in the identified concept.

The results of this preliminary experiment show the potential of our approach. The percentage of identified missing hierarchical relationships using our method is 28.26% and that of identified missing attribute relationships is as high as 69.56%. Fig. 5. Illustrates a diagrammatic example of *Scleritis and episcleritis (disorder)*, one of the identified concepts with missing hierarchical and attribute relationships. According to the assumption 1, *Scleritis and episcleritis (disorder)* is missing parents: *Scleritis (disorder)* and *Episcleritis (disorder)*. As a result of this, it is also missing the attribute relationships *Associated morphology - inflammatory morphology (morphologic abnormality)* and *Finding site - Scleral structure (body structure)*, associated with *Scleritis (disorder)*. Fig. 6. Illustrates

Table 2. Results for missing hierarchical relationships

Description	#	Percentage	# “and” Concept	# “with” Concept
Concepts for which parents were suggested (including <i>finding</i> sub-hierarchy concepts)	26	28.26%	10	16
Concepts for which parents were suggested (excluding <i>finding</i> sub-hierarchy concepts)	23	25%	10	13

Table 3. Results for missing attribute relationships

Description	#	Percentage	# “and” Concept	# “with” Concept
Concepts for which missing attribute relationships were suggested	16	69.56%	9	7

Table 4. Top three missing relationship suggestions

Concept	Suggested Parent	Suggested Attribute Relationship
Cataplexy and narcolepsy (disorder)	Cataplexy (disorder)	Finding Site - Brain structure
Miscarriage with uremia (disorder)	Uremia (disorder)	Finding site – Kidney Structure
Granulomatosis with polyangiitis (disorder)	Granulomatosis (disorder)	Associated morphology - Granulomatosis

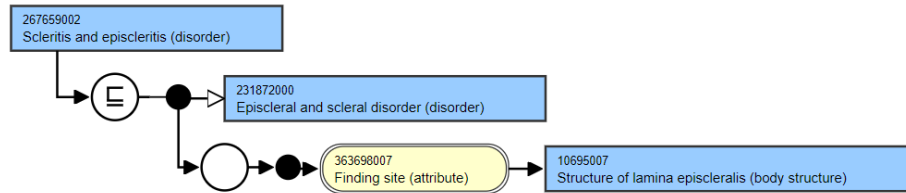


Fig. 5. Diagrammatic representation of SNOMED concept “*Scleritis and episcleritis (disorder)*” *SCTID: 267659002*”

a diagrammatic example of the suggested parent *Scleritis (disorder)* and highlights the suggested missing attribute relationships that need to be added as an additional role group to complete the modelling of *Scleritis and episcleritis (disorder)*.

Since the pilot implementation of this method has a limited scope, the following limitations are noted. Currently the method only processes FSNs containing a maximum of three words (excluding the hierarchy tag), therefore concepts containing composite-word disorder names like *Myopathy and diabetes mellitus (disorder)* (4 words), *Hepatitis A and Hepatitis B (disorder)* (5 words) are not considered in spite of being suitable candidates. Currently, the approach is not considering concepts containing “and/or” due to their complexity [19]. Lexi-

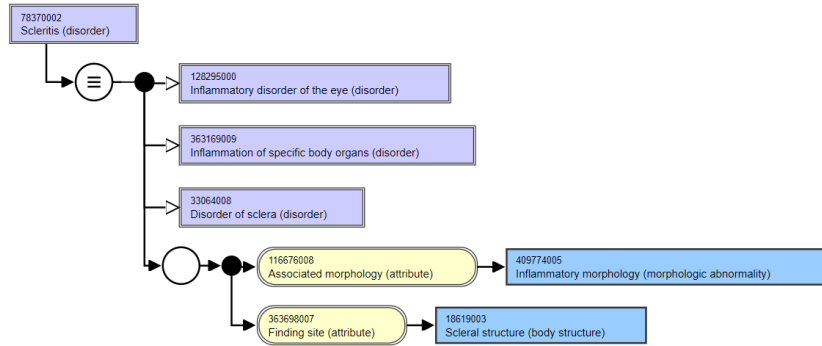


Fig. 6. Diagrammatic representation of suggested parent “*Scleritis (disorder) SCTID: 78370002*”

cal variants are generated based only on the pattern “subject + syndrome”, e.g. *osteochondrodysplasia with osteopetrosis (disorder)* is suggested a parent *osteochondrodysplasia syndrome (disorder)*. As a result other variants are neither identified as existing parents nor included in the suggested parent list. Due to lack of medical expertise on the team to verify the guidelines [1], we have for now disregarded cases where the suggested parent for a disorder belongs to the *Finding* sub-hierarchy. E.g. the suggestion of *isoimmunization (finding)* as a missing parent for the concept *pregnancy with isoimmunization (disorder)* has not been considered for further analysis. However, in spite of these limitations the method has shown promising potential and we hope to improve the accuracy of results further by working on the aforementioned limitations.

5 Conclusion and Future Work

Incomplete and inconsistent representations of CT systems cause retrieval of incorrect or partially correct result sets. Given the critical nature of medical data, the repercussions of such inaccurate results could be serious ranging from incorrect decision making in Clinical Decision Support Systems to predicting misleading trends in Population Health Management and Predictive Analytics. Thus, it is very important to implement effective QA measures for CT systems to identify any inconsistencies right at the source. In this paper, we presented a unique lexical stop-word based contextual auditing method to identify two types of inconsistencies; missing hierarchical relationships and missing attribute relationships. Employing a pilot version of this method have given promising results. The percentage of identified missing attribute relationships using our method is as high as 69.56% and that of identified missing hierarchical relationships is 28.26%. Our method has an additional asset over other QA methods that it not only identifies inconsistencies but also provides a list of potential suggestions for each identified inconsistency. Our method contributes to the improvement of a CT system in the following ways:

1. Help to produce a complete CT system by adding the suggested relationships to the CT system.
2. Ensure better extraction of inferential knowledge which is otherwise not divulged due to incomplete relationships and partially defined concepts.
3. Ensure retrieval of complete information in result sets which will facilitate informed decision making.

As future work we propose to improve our algorithm to identify composite disorder names such as *Diabetes Mellitus*. This will allow the algorithm to be applied to any FSN irrespective of its length. We plan to work on all the identified limitations. We also plan to widen the range of stop-words used in our analysis to include “of”, “due to”, “to”, etc. Finally, we will expand the technique to process FSNs containing multiple stop-words instead of a single stop-word. E.g. *Disorder due to and following burn of wrist (disorder)*.

References

1. IHTSDO, SNOMED Clinical Finding/Disorder, <https://confluence.ihtsdotools.org/pages/viewpage.action?pageId=71172245>, last accessed 2020/07/29
2. IHTSDO, SNOMED CT Browser, <https://browser.ihtsdotools.org/?>, last accessed 2020/07/30
3. PubMed Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2005-. [Table, Stopwords], <https://www.ncbi.nlm.nih.gov/books/NBK3827/table/pubmedhelp.T.stopwords/>, last accessed 2020/07/28
4. Agrawal, A.: Evaluating lexical similarity and modeling discrepancies in the procedure hierarchy of snomed ct. *BMC Medical Informatics and Decision Making* **18** (2018)
5. Agrawal, A., Elhanan, G.: Contrasting lexical similarity and formal definitions in snomed ct: Consistency and implications. *Journal of biomedical informatics* **47**, 192–8 (2014)
6. Agrawal, A., Elhanan, G., Halper, M.: Dissimilarities in the logical modeling of apparently similar concepts in snomed ct. *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2010**, 212–6 (2010)
7. Agrawal, A., Perl, Y., Chen, Y., Elhanan, G., Liu, M.: Identifying inconsistencies in snomed ct problem lists using structural indicators. *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2013**, 17–26 (2013)
8. Agrawal, A., Perl, Y., Elhanan, G.: Identifying problematic concepts in snomed ct using a lexical approach. *Studies in health technology and informatics* **192**, 773–7 (2013)
9. Agrawal, A., Perl, Y., Ochs, C., Elhanan, G.: A contextual auditing method for snomed ct concepts. *Int. J. Data Min. Bioinform.* **15**, 372–391 (2016)
10. Agrawal, A., Qazi, K.: A machine learning approach for quality assurance of snomed ct. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* pp. 792–798 (2019)
11. Agrawal, A., Qazi, K.: Detecting modeling inconsistencies in snomed ct using a machine learning technique. *Methods* (2020)

12. Agrawal, A., Revelo, P.: Analysis of the consistency in the structural modeling of snomed ct and core problem list concepts. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) pp. 292–296 (2017)
13. Bodenreider, O.: Identifying missing hierarchical relations in snomed ct from logical definitions based on the lexical features of concept names. In: ICBO/BioCreative (2016)
14. Bodenreider, O., Burgun, A., Rindflesch, T.C.: Lexically-suggested hyponymic relations among medical terms and their representation in the umls (2001)
15. Bodenreider, O., Burgun-Parenthoine, A., Rindflesch, T.C.: Assessing the consistency of a biomedical terminology through lexical knowledge. *International journal of medical informatics* **67** 1-3, 85–95 (2002)
16. Ceusters, W., Elkin, P.L., Smith, B.: Negative findings in electronic health records and biomedical ontologies: A realist approach. *International journal of medical informatics* **76** Suppl 3, S326–33 (2007)
17. Cui, L., Zhu, W., Tao, S., Case, J.T., Bodenreider, O., Zhang, G.Q.: Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in snomed ct. *Journal of the American Medical Informatics Association : JAMIA* **24**, 788 – 798 (2017)
18. Marko, K., Schulz, S., Hahn, U.: Morphosaurus–design and evaluation of an interlingua-based, cross-language document retrieval engine for the medical domain. *Methods of information in medicine* **44** 4, 537–45 (2005)
19. Mendonça, E.A., Cimino, J.J., Campbell, K.E., Spackman, K.A.: Reproducibility of interpreting "and" and "or" in terminology systems. *Proceedings. AMIA Symposium* pp. 790–4 (1998)
20. Pacheco, E.J., Stenzhorn, H., Nohama, P., Paetzold, J., Schulz, S.: Detecting under-specification in snomed ct concept definitions through natural language processing. *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2009**, 492–6 (2009)
21. Rector, A.L., Iannone, L.: Lexically suggest, logically define: Quality assurance of the use of qualifiers and expected results of post-coordination in snomed ct. *Journal of biomedical informatics* **45** 2, 199–209 (2012)
22. Schulz, S., Martínez-Costa, C., Miñarro-Giménez, J.A.: Lexical ambiguity in snomed ct. In: JOWO (2017)