Annotation of Chest Radiology Reports for Indexing and Retrieval

Dina Demner-Fushman, Sonya E. Shooshan, Laritza Rodriguez, Sameer Antani, and George R. Thoma

Lister Hill National Center for Biomedical Communications,
U.S. National Library of Medicine, Bethesda, MD
{ddemner,sshooshan,rodriguezlm2,santani,gthoma}@mail.nih.gov

Abstract. Annotation of MEDLINE citations with controlled vocabulary terms improves the quality of retrieval results. Due to variety in descriptions of similar clinical phenomena and abundance of negation and uncertainty, annotation of clinical radiology reports for subsequent indexing and retrieval with a search engine is even more important. Provided with an opportunity to add about 4,000 radiology reports to collections indexed with NLM image retrieval engine Open-i, we needed to assure good retrieval quality. To accomplish this, we explored automatic and manual approaches to annotation, as well as developed a small controlled vocabulary of chest x-ray indexing terms and guidelines for manual annotation. Manual annotation captured the most salient findings in the reports and normalized the sparse distinct descriptions of similar findings to one controlled vocabulary term. This paper presents the vocabulary and the manual annotation process, as well as an evaluation of the automatic annotation of the reports.

Keywords: Information Storage and Retrieval; Vocabulary, Controlled; Radiology

1 Introduction

Manual annotation of MEDLINE® citations with controlled vocabulary terms (Medical Subject Headings (MeSH®) [1][2] indexing) significantly improves the quality of the retrieval results for Boolean and ranking search engines [3] [4] [5]. The reasons for the improvements are twofold: firstly, the indexing is based on the full text of the paper, rather than the title and the abstract that are often the only texts available for automatic indexing using search engines; secondly, the indexers summarize the "aboutness" and the most salient aspects of the paper, both hard tasks to automate, particularly given only the abstracts of the papers.

Significant advances in automated indexing of MEDLINE citations have allowed National Library of Medicine (NLM®) Medical Text Indexer (MTI) system to become a first line indexing tool for selected journals, which allows NLM indexers to revise automatic indexing results, rather than indexing these journals manually[6]. Without testing, it is hard to predict if MTI or other automatic approaches will be successful in preparing collections of clinical notes for indexing and retrieval.

Provided with an opportunity to add about 4,000 publicly available radiology reports offered by Indiana University to the collection of over 500,000 biomedical articles indexed with NLM image retrieval engine Open-i [7], we needed to annotate the reports for most salient findings to ensure better retrieval results. We first researched the available terminologies that could be used for annotation of the reports: RadLex [8][9], MeSH, and the Unified Medical Language System® (UMLS®)[10][11]. Our preference was to use MeSH, to be consistent with MeSH indexing of the enriched biomedical citations in Open-i. We expected RadLex to better capture the terminology used in these reports and anticipated that we would need to augment MeSH with RadLex terms. To infer what search terms might be used by the Open-i users, we studied radiology textbooks [12][13] and topics in the ImageCLEF evaluations [14][15]. We also mapped the reports to the UMLS using MetaMap[16], and analyzed the frequently occurring terms. This preliminary analysis confirmed that we need to combine terms from MeSH and RadLex, as well as establish term similarity sets to allow retrieving reports that describe similar radiological findings using different terms. Although our task was much simpler than building the UMLS[10] or the Open Biological and Biomedical Ontology Foundry(OBO Foundry)[17], the principles developed for building these ontologies guided our approach to the task. Building our vocabulary was most similar to building or borrowing a taxonomy for indexing electronic medical records[18][19]: MeSH and RadLex served as the basis for our vocabulary and we used clinical data as a filter to capture only the terms needed for our task.

The goal of our annotation was to capture the most salient findings described in the reports (akin to major MeSH descriptors in MEDLINE citations) and to normalize the sparse distinct descriptions of the same findings to a controlled vocabulary term. Therefore, our controlled vocabulary is minimal and constructed primarily "bottom-up" to capture the information in 4,000 chest radiology reports. In this paper, we present the controlled vocabulary and guidelines for manual annotation of chest radiology reports for subsequent indexing and retrieval in our search engine, Open-i. We also evaluate the results of automatic annotation of the reports using MTI and a UMLS-based tool.

2 Methods

Our overall goal is to facilitate retrieval of relevant radiology reports with a search engine. To achieve this goal, we assign controlled vocabulary terms to the reports using three methods: 1) MTI that assigns MeSH terms; 2) SGindexer that uses MetaMap to extract asserted UMLS concepts in the Disorders and Procedures semantic groups[20]; and 3) manual annotation. Our preliminary study indicated we could not use any single existing terminology as the basis for manual annotation: the manual annotation approach required building a controlled vocabulary and developing guidelines for annotation.

2.1 Building a controlled vocabulary for manual annotation

In building the vocabulary, we followed the steps outlined by Zaharee[21]:

Determine Scope: we defined the scope as a minimal set of terms needed to describe the major radiologic findings in the reports available to us.

Identify Sources: we chose MeSH and RadLex as the basis taxonomies and turned to the UMLS, Felson's radiology textbook and, most importantly, the Indiana University radiology reports for additional information.

Plan for Maintenance: For the terms that we borrowed from MeSH and RadLex, our plan is to stay current with updates to these sources.

Gather Terms: We combined the top-down and bottom-up methods described by Zaharee. We started with a top-down vocabulary built using the diseases, anatomical sites, imaging observations, objects, and qualifiers found in Felson's radiology textbook. We then annotated 100 longest reports and identified terms that were missing in our vocabulary, as well as the terms that were not used. We removed the unused terms and added the missing terms to the vocabulary. Several missing terms encountered later during the annotation and validation processes were added to the vocabulary. All terms were added to the vocabulary based on the consensus opinion of three annotators described in the next section.

Categorize Terms: We mapped the terms to the preferred MeSH Descriptor names, where possible. After consulting with radiology textbooks and the existing ontologies, we grouped related terms into similarity sets, for example, the similarity set for the preferred term Osteophyte includes: bony exostosis, external hyperostosis, bone spur, exostoses, osteophyma, osteophytosis, syndesmophyte, and spurring. We assigned terms to five categories: Diseases, Anatomy, Objects, Signs, and Attributes. We mapped the terms to 2014 MeSH and RadLex version 3.10. Term similarity sets were automatically mapped to MeSH and RadLex using exact string matches between each term in a similarity set and terms in RadLex 3.10 and the 2014 MeSH Tree Structure file. If a unique identifier was found, it was assigned to the term. These mappings were verified manually to exclude spurious partial matches. We established parent/child relations between the terms in the Diseases, Anatomy, Objects and Signs categories. The Attributes category, which is akin to the MeSH qualifiers, was further sub-categorized into: degree, descriptive, spatial, size, and type groups.

Manage Terms. Visualize Terms. Export Terms: Due to the small size and simple structure of our vocabulary, it was possible to manage, visualize and export it using an Excel spreadsheet.

Review/Validate: We reviewed and validated the terms in the course of the annotation process and during evaluation of the automatic annotation.

Post to a Registry or Data Warehouse: The controlled vocabulary is available through Open-i services.

2.2 Annotating radiology reports

We annotated 3,955 de-identified chest radiology reports provided by Drs. Kohli and Rosenman, Indiana University. The reports are formatted according to standards[22] and contain the traditional *Clinical Information*, *Comparison*, *Findings*, and *Impression* sections. We used the *Findings* and *Impression* sections in all our annotations, with more weight for the *Impression* section.

MTI annotation MTI combines UMLS concepts found by MetaMap with terms from related citations recommended by the PubMed Related Citations algorithm[23], and then restricts the terms to MeSH, clusters, and ranks them[6]. We formatted the reports as MEDLINE citations: the *Impression* section was placed in the title field because MTI gives more weight to the terms found in the title. The *Findings* section was placed in the abstract field. We then submitted the reports to MTI using the settings provided in the MeSH on Demand[1] application powered by MTI.

SGindexer The *Impression* and *Findings* section were submitted to MetaMap with default settings. The output was restricted to semantic types: Congenital Abnormality, Acquired Abnormality, Injury or Poisoning, Pathologic Function, Disease or Syndrome, Anatomical Abnormality, Neoplastic Process, and Sign or Symptom in the semantic group Disorders, as well as Therapeutic or Preventive Procedure and Diagnostic Procedure in the semantic group Procedures. Only concepts that were definitely not negated according to NegEx[24] implementation of MetaMap were selected for annotation.

Manual annotation Two annotators: an experienced clinician (LR) and a medical librarian (SES), both trained in medical informatics and experienced in medical document annotation, independently annotated each report using Excel spreadsheets. We implemented a comparison program in Java to find disagreements between the annotators. The program compared annotations for each document and identified missing terms and attributes. The output of the program indicated which terms and attributes were missing for each annotator for a given report. The annotators then reconciled the disagreements. In the rare cases when the annotators could not come to an agreement, the remaining disagreements were adjudicated with the help of the third annotator with clinical and medical informatics experience (DDF).

In the first annotation pass, we annotated the reports at a coarse level as normal chest or abnormal. If the report mentioned the deficient quality of the image, it was flagged as such. The second annotation pass excluded the normal reports and focused on identifying major findings that characterized the images discussed in the reports. To keep the size of our vocabulary small, we

[1] `http://www.nlm.nih.gov/mesh/MeSHonDemand.html`

decided to post-coordinate the atomic terms in the vocabulary akin to the descriptor/qualifier combinations in MeSH. For example, instead of adding the UMLS concept *Left lower lobe pneumonia* to our vocabulary, we combine the terms as follows: *Pneumonia/lower lobe/left.*

Manual annotation guidelines Our guidelines follow the general NLM indexing guidelines[25]. To avoid bias, we chose to perform unassisted annotation. Each annotation corresponds to a single major radiological finding. Terms in the Diseases, Objects and Signs categories can serve only as descriptors and are placed in the first position in each annotation. The remaining terms in each annotation serve as qualifiers and are in the Anatomy or Attributes category. Terms in the Anatomy category can serve as descriptors, if needed, for example, *Lung/hyperinflation.* The qualifiers are added to the annotation in the following order: anatomical site(s) in descending order of granularity, for example, *Opacity/lung/base*; spatial attribute(s), for example, *Opacity/lung/base/right*; and size or degree, for example, *Opacity/lung/base/right/round/multiple/small* or *Opacity/lung/base/right/mild.*

The primary difference between annotating radiology reports and MEDLINE citations is in the abundance of negation and uncertainty (hedging) in the reports. Whereas it is easy to decide that negated terms should not be annotated, the degrees of uncertainty were harder to capture in the guidelines. After annotating the first 100 reports, we determined that in some cases the hedging term indicated the pathology was present and the uncertainty was in the specifics, therefore the hedging terms could be ignored. For example, in the *Impression*: `The changes are compatible with known diagnosis of sarcoidosis`, *compatible with* is ignored, and the report is annotated with the term *Sarcoidosis.* The hedging terms that we decided to ignore to capture salient findings are: *probably*, *consistent with*, *likely*, *compatible*, and *most suggestive.* Most cases of hedging, however, were not annotated.

Finally, we had to consider the consequences of the fairly rigorous de-identification process. In some cases, we were not able to determine whether the part of a term that was not de-identified belonged to a finding, for example in `XXXX change in the XXXX XXXX alignment of the scapula XXXX and spine`. In these cases, nothing was annotated. If the de-identified part of a term could be reliably inferred and the term was an important finding, we annotated the reports with the term. For example, we assigned *Costophrenic Angle/left/blunted* to the *Impression*: `Stable blunting of the left costophrenic XXXX which may represent persistent left pleural effusion versus pleural scarring`. Because of the uncertainty about *pleural effusion* or *scarring*, this was the only annotation for the report.

2.3 Evaluation

Two annotators (LR and SES) manually reviewed terms assigned to 3,955 reports by MTI and SGindexer. In the process, they also validated the manual annotations and the controlled vocabulary. The automatically assigned terms were judged 1) correct, 2) neutral, 3) somewhat incorrect, and 4) incorrect. Terms were judged correct if a major finding was correctly identified. Terms were judged neutral if the annotation was correct, but the term described trivial anatomy or findings, such as, *rib*, *surgery* or *human*. Terms were judged somewhat incorrect if a part of the term that was captured did not have an appropriate sense in any of the source vocabularies. For example, reports often mention *bronchovascular crowding*, which is not a term in any of the sources, however *crowding* is a UMLS concept with semantic type social behavior [$C0010383$], as well as a synonym of *tooth crowding* [C0040433] and *dental crowding* [$C1847525$]; and was, therefore, suggested by SGindexer. When captured by itself, crowding does not appear to be incorrect, however the CUIs reveal that it maps to wrong concepts. Terms were judged incorrect if automatic annotation captured a term that was negated or was not stated in the report. Each annotator annotated over 2,000 reports, with 500 in common, so that we could compute inter-annotator agreement. We computed the agreement using Cohen's kappa[26].

In addition to the manual evaluation of extracted terms, we computed precision and recall for MTI and SGindexer, using as reference set the union of the manually assigned terms and the manually-identified correct terms found by both tools. We ignored the neutral terms when computing recall and precision. We judged the somewhat incorrect and incorrect terms to be false positives. We computed *Recall* as the ratio of correct terms found by the tool to the number of terms in the reference set: $tp/(tp + fn)$. We computed *Precision* as the number of correct terms assigned by the tool divided by the sum of the correct, somewhat incorrect and incorrect terms assigned by the tool: $tp/(tp + fp)$. We computed both the macro-average recall and precision that gives equal weight to each term, and the micro-average that gives equal weight to each per-document term assignment.

3 Results

In the first pass of manual annotation, $2,314$ of $3,955$ documents were labeled abnormal. In the second pass, $6,519$ terms were assigned to $2,314$ abnormal reports, ranging from 1 to 13 terms per report, and close to 3 on average. *Deficient Quality* flags were assigned to 91 images. The most detailed terms (with a descriptor term and seven qualifiers) assigned to a report were:

- Opacity/lung/upper lobe/bilateral/reticular/round/multiple/chronic
- Opacity/lung/base/bilateral/scattered/focal/patchy/multiple
- Opacity/lung/bilateral/interstitial/diffuse/reticular/round/severe
- Opacity/lung/middle lobe/bilateral/interstitial/round/small/mild

We assigned 861 terms without attributes, primarily in the Diseases category, for example, *Osteoporosis*, and some in the Objects category, for example, *Catheters, Indwelling*. The final controlled vocabulary contains 47 anatomy terms, 62 diseases terms, 10 objects, 17 signs, and 41 attributes (counting the *Deficient Quality flag* that maps to the RadLex term *limited quality* $RID13$.)

The similarity sets of these 177 preferred terms contain the total of 487 terms. The preferred term *Implanted Medical Device* has the largest similarity set: *cardioverter defibrillators, implantable*; *automatic internal defibrillators*; *automatic defibrillator*; *defibrillators, internal*; *icd*; *stimulator*; *Defibrillators, Implantable*; *Atrial septal occluder*; *Aortic valve prosthesis*. The automatic mappings of the terms to MeSH, RadLex and UMLS revealed that 55 terms had no exact matches in RadLex, 88 had no exact matches in MeSH, and 51 could not be mapped to the UMLS using the same MetaMap settings and restrictions to semantic types that we used for SGindexer. Table 1 shows the distribution of the terms in the sources and the overlaps between the three sources.

Table 1. Distribution of radiology report indexing terms in MeSH, RadLex, and UMLS.

Category	Present in all	UMLS & RadLex	UMLS & MeSH	MeSH only	RadLex only	Total
Diseases	44	8	9	1	-	62
Signs	1	11	4	-	1	17
Anatomy	34	12	-	-	1	47
Objects	7	2	1	-	-	10
Attributes	-	39	-	-	2	41
Total	86	72	14	1	4	177

Only 10 terms could not be mapped to any source directly. These terms have multiple partial or even full matches in MeSH, UMLS and RadLex, as shown in Table 2. The table also shows reasons for missing the mappings in the automatic process. During the manual review, we were able to find an equivalent MeSH or RadLex term for each of the preferred terms, mapping 101 terms to MeSH and 76 to RadLex.

3.1 Automated indexing results

MTI has assigned a total of 11,923 non-neutral terms to 3,955 documents. The SGindexer has assigned 6,620 non-neutral terms. The reference set contained 11,987 terms. The recall and precision for both tools are shown in Table 3 .

The inter-annotator agreement for the 500 reports annotated in common was good, with $\kappa = 0.84$. The best agreement was for the correct and incorrect judgments ($\kappa = 0.86$ and 0.92, respectively).

In the validation phase, when the automatically assigned terms were compared to the manual terms, the annotators assigned 342 additional annotations (mostly describing chronic bone and skeletal changes). They also added one synonym (*TIPS*) to the set of *Catheters, Indwelling* one synonym (significantly) to

Table 2. Terms without an exact automatic match and their coverage in MeSH, RadLex, and UMLS.

Term	MeSH		RadLex		UMLS		Example term (source)
	#	reason missed	#	reason missed	#	reason missed	
expansile lesion	1	Too general	2	Post-coordination	0	Too specific	Expansile Bone Lesions (MeSH)
cardiophrenic angle	0	No match	1	Variant	3	Too general	cardiophrenic sulcus (RadLex)
lucency	0	No match	1	Wrong POS	15	Too general	lucent (RadLex)
shift	1	Too general	1	Wrong POS	1	Wrong ST	Shifted (RadLex)
supracardiac	0	No match	1	Too specific	1	Wrong ST	cardiac region (RadLex)
multilobar	0	No match	1	Variant	3	Too general	multilobulated (RadLex)
retrocardiac	0	No match	1	Too specific	2	Too general	cardiac region(RadLex)
borderline	0	No match	1	Synonym	1	Wrong ST	insignificant (RadLex)
paratracheal	0	No match	5	Too general	1	Wrong ST	paratracheal lymph node (RadLex)
streaky	0	No match	1	Synonym	0	No match	linear (RadLex)

stands for the number of related concepts in MeSH, UMLS and RadLex; Too specific means our term is more specific than the terms in the sources; Too general means our term is more general than the terms in the sources; Post-coordination indicates it might be possible to generate our term post-coordinating the terms in the source; Wrong ST means we missed an exact match due to exclusion of the UMLS semantic types. Wrong POS means we missed a match due to part-of-speech mismatch, for example, shift/shifted.

Table 3. Micro- and Macro-averaged recall and precision (in %) for MTI and SGindexer

Tool	Macro-averaging		Micro-averaging (per document)		Totals of non-neutral terms		
	Recall	Precision	Recall	Precision	True positive	False positive	False negative
MTI	28.7	28.9	16.5	23.5	3445	8478	8542
SGindexer	40.5	73.3	38.6	68.1	4854	1766	7133

the set of *severe* and two synonyms (syndesmophyte and spurring) to the set of *Osteophyte*.

4 Discussion

Previous research indicated that searching the full text of radiology reports for a radiological finding will retrieve many documents in which the term is mentioned because it was provided as reason for examination or because the patient does not have this problem[24]. Clearly, we needed to provide a mechanism similar to MeSH indexing of MEDLINE citations to help searchers find reports that assert the presence of the radiologic findings of interest to them. We were uncertain if the tools available to us for automatic annotation could reliably identify such terms and if manual annotation was necessary. Our analysis shows that the two automatic tools evaluated in our study are not yet ready to be used for automatic annotation of clinical text for subsequent indexing and retrieval in search engines. Whereas SGindexer has somewhat better precision due to handling negations and uncertainty, and better recall because it relies on the full UMLS Metathesaurus, both tools still have relatively low recall. We attribute MTI?s low precision (less than half of the MTI precision for MEDLINE citations) to the fact that it is not handling negation. The correct terms suggested by MTI, however, provided insights and were often used to augment our manual annotation. We are also not surprised by MTI?s low recall: restriction to MeSH prevented it from finding terms at the granularity level we chose for the reports. We believe that if MTI will take into account negation and hedging, it will be a valuable tool for assisted indexing of clinical reports. The poor SGindexer recall could be explained by the absence of many RadLex terms in the UMLS and by the fact that we used only a few semantic types. For example, the MeSH only column in Table 1 should be empty, as MeSH is included in the UMLS. Restricting our mappings to only a few semantic types to have better precision, we naturally reduced our recall. For example, our vocabulary term *hypovolemia* in the Diseases category was mapped only to MeSH because its semantic type is finding; SGindexer, therefore, missed it. Similarly to MTI, the primary reasons for somewhat lower than expected precision of extraction of the terms in the semantic group Disorders using MetaMap is due to errors in detecting negation and uncertainty. As with MTI, we believe that with improvements to the negation algorithm and better selection of allowed semantic types, SGindexer could be used for assisted annotation of clinical reports.

Not surprisingly, our manual annotation provided a different view of the reports, compared to the automatic annotation. This, however, was a time-consuming process. Recognizing a normal chest report took approximately 5 minutes for each. Annotating each abnormal report took 10 to 20 minutes, depending on the length and complexity of the text. The reconciliation process took approximately 40 hours, during which we not only reconciled the differences, but also worked on the controlled vocabulary, particularly, on the similarity sets.

Our vocabulary borrows 101 preferred names from MeSH and 76 from RadLex. Similarly to Bekhuis et al.[27], we found that terminologies vary and the union of the two sources was the best basis for our vocabulary. In ten cases, we could not match the terms to the sources exactly because we chose to annotate with more general terms and post-coordination, rather than having a sparse vocabulary of more exact terms. For example, our term thoracic aorta is more general than descending thoracic aorta (RID35844) and segment of thoracic aorta (RID35843) in RadLex, however it is covered by the MeSH term *Aorta, Thoracic*. The majority of the Disease and Anatomy terms is covered by all three sources. As anticipated, the Attributes are not covered in MeSH. It was surprising however, that most Attributes available in RadLex are also covered in the UMLS.

To further reduce sparseness, our similarity sets mostly cover more than one concept in the three sources. For example, the set for Granulomatous disease maps to one RadLex concept RID34787 (Granulomatous disease) and two UMLS concepts: C0740451 (Granulomatous disorder) and C1610637 (Granulomatous infection). Our complete set contains the following terms: Evidence of prior granulomatous disease; old granulomatous disease; Granulomatous disorder; Evidence of previous granulomatous infection. Interestingly, we could not assign any MeSH terms to this concept, because we felt that the term chronic granulomatous disease that is available in MeSH is not the same as evidence of prior disease, and because its parent, Phagocyte Bactericidal Dysfunction, is too general for our purposes. We were somewhat surprised that we could not rely on automatic matching of the terms in radiology reports to RadLex. Table 2 shows that finding equivalents for some terms required inference. We hope that the *Imaging Observation* section of RadLex will be expanded in the future.

Although manual annotation seems useful for Open-i searches as shown in Figure 1, we plan to rigorously compare the relevancy of the retrieval results for the three annotation methods in the future. We also plan to continue exploring automatic annotation using all UMLS semantic types and adding RadLex to the vocabularies.

5 Conclusions

Our overall goal is to facilitate retrieval of relevant radiology reports with a search engine. In this study we have demonstrated that a small controlled vo-

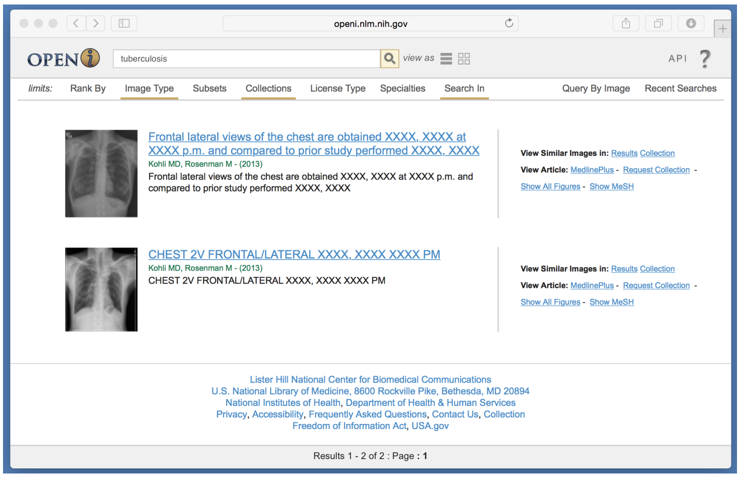

Fig. 1. Searching Open-i for *Tuberculosis* in manual annotation only delivers 2 results shown in the figure. Searching in the full text of the reports retrieves 58 reports, most of which are not relevant. For example, `Impression: Normal chest No evidence of tuberculosis.` Note that the *XXXX* patterns in the reports are the consequences of the de-identification process.

cabulary and post-coordination of terms could be used to capture salient findings in chest radiology reports for subsequent indexing and retrieval in a search engine. We used a combination of MeSH, and RadLex terms as a basis for our vocabulary. Although these two sources provide terms exactly or approximately equivalent to all findings in the radiology reports, the language of the reports differs from both sources and manual approximate matching was needed to cover all findings.

We further evaluated if the existing tools could be used to automatically annotate the reports and concluded that the tools need to reliably recognize negation and hedging to render accurate results. In addition, vocabularies need more synonymy and better coverage to enable direct matches. Automatic UMLS-based mappings that took into account negation achieved 40% recall at 73% precision. These results indicate that the automatic annotation tools could assist manual annotation of clinical text for indexing and retrieval with search engines.

The Indiana University radiology collection, the vocabulary and the guidelines are provided through utlhttp://openi.nlm.nih.gov/contactus.php

Acknowledgments

This work was supported by the intramural research program of the U. S. National Library of Medicine, National Institutes of Health.

References

1. Rogers FB. Medical subject headings. Bull Med Libr Assoc. Jan;51:114-6. (1963)
2. MeSH . Available from: `http://www.nlm.nih.gov/mesh/meshhome.html`
3. Haynes RB, Wilczynski N, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. J Am Med Inform Assoc. Nov-Dec;1(6):447-58.(1994)
4. Hersh W, Voorhees E. TREC genomics special issue overview. Inf. Retr. Feb; 12 (1): 1-15.(2009)
5. Darmoni SJ, Soualmia LF, Letord C, Jaulent MC, Griffon N, Thirion B, Nvol A. Improving information retrieval using Medical Subject Headings Concepts: a test case on rare and chronic diseases. J Med Libr Assoc. Jul;100(3):176-83. (2012)
6. Mork JG, Jimeno-Yepes AJ, Aronson AR. The NLM Medical Text Indexer System for Indexing Biomedical Literature. BioASQ Workshop, Valencia, Spain, September 27. (2013)
7. Demner-Fushman D, Antani S, Simpson MS, Thoma GR. Design and development of a multimodal biomedical information retrieval system. JCSE v.6, no.2: 68-177. (2012)
8. Langlotz CP. RadLex: a new method for indexing online educational materials. Radiographics. Nov-Dec;26(6):1595-7. (2006)
9. RadLex. Available from: `http://www.radlex.org/`
10. Humphreys BL, Lindberg DAB. Building the unified medical language system. In Proc Annu Symp Comput Appl Med Care, pp. 475-80. (1989)
11. UMLS. Available from: `https://uts.nlm.nih.gov/home.html`
12. Goodman LR, Felson B. Felson's principles of chest roentgenology?: a programmed text. Philadelphia: Saunders. (1999)
13. Daffner RH. Clinical radiology: the essentials, 2nd ed. Baltimore: Williams & Wilkins,. (1999)
14. Hersh WR, Mller H, Jensen JR, Yang J, Gorman PN, Ruch P. Advancing biomedical image retrieval: development and analysis of a test collection. J Am Med Inform Assoc. Sep-Oct;13(5):488-96. (2006)
15. The ImageCLEF campaign. Available from: `http://www.imageclef.org/2013/medical`
16. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc. May-Jun;17(3):229-36. (2010)
17. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology, 25(11), 1251-1255. (2007)
18. Kuranz J, Gilles B. Indexing electronic medical records using a taxonomy. Bulletin of the American Society for Information Science and Technology 39, no. 2: 30-33. (2013)
19. Call B. Indexing electronic medical records using a taxonomy. In Proceedings of the 2013 international workshop on Data management & analytics for healthcare, pp. 5-8. ACM. (2013)
20. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Stud Health Technol Inform (Proc Medinfo);84(Pt 1):216-220. (2001)
21. Zaharee M. Building controlled vocabularies for metadata harmonization. Bulletin of the American Society for Information Science and Technology 39, no. 2: 39-42. (2013)

22. RSNA radiology reporting initiative. Available from: http://www.radreport.org/template/0000102
23. Lin J, Wilbur WJ. PubMed related articles: a probabilistic topic-based model for content similarity. BMC bioinformatics, 8(1), 423. (2007)
24. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform. Oct;34(5):301-10. (2001)
25. MEDLINE Indexing Online Training Course. Available from: http://www.nlm.nih.gov/bsd/indexing/training/USE_010.htm
26. Siegel S, Castellan N. Nonparametric statistics for the behavioral sciences. 2nd Ed. New York: McGraw-Hill. (1988)
27. Bekhuis T, Demner-Fushman D, Crowley RS. Comparative effectiveness research designs: an analysis of terms and coverage in Medical Subject Headings (MeSH) and Emtree. J Med Libr Assoc. Apr;101(2):92-100. (2013)