Original Article

Healthc Inform Res. 2015 January;21(1):35-42.
http://dx.doi.org/10.4258/hir.2015.21.1.35
pISSN 2093-3681 • eISSN 2093-369X

HIR
Healthcare Informatics Research

Challenges and Practical Approaches with Word Sense Disambiguation of Acronyms and Abbreviations in the Clinical Domain

Sungrim Moon, PhD[1], Bridget McInnes, PhD[2], Genevieve B. Melton, MD, MA[3,4]

[1]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX; [2]Department of Computer Science, Virginia Commonwealth University, Richmond, VA; [3]Institute for Health Informatics and [4]Department of Surgery, University of Minnesota, Minneapolis, MN, USA

Objectives: Although acronyms and abbreviations in clinical text are used widely on a daily basis, relatively little research has focused upon word sense disambiguation (WSD) of acronyms and abbreviations in the healthcare domain. Since clinical notes have distinctive characteristics, it is unclear whether techniques effective for acronym and abbreviation WSD from biomedical literature are sufficient. **Methods:** The authors discuss feature selection for automated techniques and challenges with WSD of acronyms and abbreviations in the clinical domain. **Results:** There are significant challenges associated with the informal nature of clinical text, such as typographical errors and incomplete sentences; difficulty with insufficient clinical resources, such as clinical sense inventories; and obstacles with privacy and security for conducting research with clinical text. Although we anticipated that using sophisticated techniques, such as biomedical terminologies, semantic types, part-of-speech, and language modeling, would be needed for feature selection with automated machine learning approaches, we found instead that simple techniques, such as bag-of-words, were quite effective in many cases. Factors, such as majority sense prevalence and the degree of separateness between sense meanings, were also important considerations. **Conclusions:** The first lesson is that a comprehensive understanding of the unique characteristics of clinical text is important for automatic acronym and abbreviation WSD. The second lesson learned is that investigators may find that using simple approaches is an effective starting point for these tasks. Finally, similar to other WSD tasks, an understanding of baseline majority sense rates and separateness between senses is important. Further studies and practical solutions are needed to better address these issues.

Keywords: Abbreviations as Topic, Medical Records, Natural Language Processing, Artificial Intelligence, Automated Pattern Recognition

Submitted: December 23, 2014
Accepted: January 19, 2015

Corresponding Author
Sungrim Moon, PhD
School of Biomedical Informatics, 7000 Fannin St. Suite 870, Houston 77030, TX, USA. Tel: +1-713-500-3993, Fax: +1-713-500-3907, E-mail: Sungrim.Moon@gmail.com

I. Introduction

The use of acronyms and abbreviations in both the biomedical and clinical domains is pervasive and increasing rapidly [1-5]. In clinical medicine, one of the main impetuses for this increasing use of acronyms and abbreviations is the fast-growing adoption of Electronic Health Record (EHR) systems resulting in the proliferation of electronic clinical documents. In addition to electronic clinical notes that are traditionally created by dictation and transcription, many clinical notes are now manually created within a time-constrained clinical work environment in which clinicians type or enter notes fully or use a semi-structured or templated document entry system. This often results in the use of shortened word forms that are frequently ambiguous and present a problem for subsequent information retrieval from EHR systems and potentially can lead to patient safety issues [6-8].

Development and improvement of automated techniques to resolve the sense of acronyms and abbreviations in clinical text is an important challenge for medical natural language processing (NLP), and it is considered an essential component for automated medical NLP systems [3]. Acronym and abbreviation sense resolution is considered a special case of word sense disambiguation (WSD) [9-11]. While interpreting the specific meaning of acronyms and abbreviations within a sentence is often easy for a human reader, this process is non-trivial for a machine [10,11]. In general English, studies have demonstrated that humans can properly resolve the meaning of most acronyms and abbreviations even if only given a very limited context of five words with the acronym or abbreviations in the central position [12,13]. Machine learning techniques have been used extensively to address the problem of automatic WSD in the biomedical and general English domains. Supervised methods have also been demonstrated in analogous studies to be potentially well suited for clinical acronym disambiguation [1,4,14,15].

Although acronym and abbreviation WSD has been studied extensively in relation to biomedical literature, relatively little research has been devoted to WSD of acronyms and abbreviations within clinical notes. In biomedical literature, typically the first instance of a short form for the acronym or abbreviation occurs with long form as a parenthetical expression or vice versa (e.g., mucosal ulcerative colitis) [11]. Because clinical notes are informal in nature, the association of the long form and short form in clinical text is rarely observed [5,15]. Moreover, the development of automated approaches for the disambiguation of acronyms and abbreviations in clinical text is complicated by legitimate issues of patient confidentiality and privacy which represent significant barriers for sharing clinical notes for research purposes [3,16].

Up to now, there has been limited utilization of clinical document characteristics and clinical knowledge resources for automated techniques addressing clinical acronym and abbreviation WSD. This article discusses issues encountered in acronym and abbreviation WSD using clinical notes from a tertiary healthcare institution. We then propose possible practical solutions to these problems based on our experiences to date.

II. Methods

Clinical documents from the University of Minnesota-affiliated Fairview Health Services, including the University of Minnesota Medical Center and three additional hospitals in the Minneapolis metropolitan area, from a 5-year period were used as our corpus for this study. This corpus was composed primarily of dictated clinical notes which were subsequently manually transcribed and stored in electronic format. These documents included admission history and physical examinations, consultations, and discharge summaries.

Acronyms and abbreviations were identified by using a set of heuristic rules with regular expressions and Perl scripts. An acronym or abbreviation for this study was defined as a token consisting of capital letters, numbers, and symbols (period, comma, colon, or semicolon). If this word token had more than 500 occurrences in the corpus, it was considered a potentially clinically significant acronym or abbreviation. After these acronyms and abbreviations were detected, 500 random occurrences were selected within the corpus, along with the surrounding previous 7 tokens and subsequent 7 tokens. These extracted instances were presented to two physicians who participated in this study to manually annotate for the senses of acronyms and abbreviations. The annotated sense of the acronym or abbreviation was then used as the reference standard.

Our goal was to obtain the optimal feature selection method for automated machine learning (ML) techniques to disambiguate clinical acronyms and abbreviations. To do this, we extracted a number of potentially predictive features by utilizing resources and techniques from a number of disciplines including biomedical NLP, computational linguistics, statistics and clinical practice. The feature types that were explored included 1) bag of words (BoW), which is defined as the set of surrounding non-normalized word tokens of the targeted acronym or abbreviation; 2) biomedical concepts via the Unified Medical Language System (UMLS) concept unique identifier (CUI); 3) biomedical semantic information (the UMLS semantic type of each concept); 4) linguistic information with parts of speech; 5) term frequency with other statistical information; as well as 6) heuristic clinical note structure and title/section information. We implemented the BoW approach using the set of surrounding 14 non-normalized word tokens and excluded English stop words that do not hold significant semantic information. We used a standard list of 57 English stop words [17]. MetaMap [18] was utilized to obtain both UMLS CUIs and UMLS semantic types for the targeted 14 surrounding words with the

http://dx.doi.org/10.4258/hir.2015.21.1.35

targeted acronym or abbreviation. MetaMap automatically processes the text and performs normalization and stemming as well. We also grouped semantic information using the previously described 15 semantic groupings of UMLS semantic types [19] as a set of features with UMLS semantic type information. Lastly, we extracted section names using a combination of heuristic rules with regular expressions and then ensured proper grouping of equivalent sections using manual classification by a physician. For the purposes of this pilot study, linguistic and statistical features were not included in the initial set of experiments.

Several supervised ML algorithms through the Weka data mining package [20] were used on each of the feature sets. We explored the application of naïve Bayes, support vector machines (SVM), and decision trees. In Weka, the specific algorithms are NaiveBayes, LibSVM, and J48, respectively. For our evaluation, we relied on the 10-fold cross-validation functionality implemented in Weka to assess performance for each abbreviation or acronym set of samples. Performance for each set of features and ML algorithm was measured in terms of precision, recall, and F-measure.

III. Results

As part of the 'lessons learned', we present the challenges faced to date, as well as practical learning from ongoing experiments in order to develop an effective medical NLP module for WSD of acronyms and abbreviations from clinical documents.

1. Practical Difficulties with Using Clinical Text

In the initial phases of our research, some of the key challenges that we encountered originated from several broad areas, including 1) variation in format, structure, and proper use of language in clinical texts including the common use of sentence fragments; 2) a shortage of resources, tools, and knowledge based on clinical notes; and 3) privacy issues regarding the use of protected health information.

1) Challenge 1: Language, structure, and formatting

Because the primary function of clinical notes is to record medical information conveyed between clinicians as a form of communication and documentation, these notes are not created with the intention of re-use or for the purpose of helping researchers perform automated tasks, such as WSD of abbreviations and acronyms. Therefore, one primary difficulty encountered in our work was the lack of formal structure in notes and format standardization of documents in EHR systems within the error-prone clinical environment. For example, there are portions of many clinical notes with extraneous text, which is not helpful for WSD research purposes. This includes formatting at the beginning and end of documents, extra white space, and the informal use of tables. Sometimes these are institution-specific formatting issues particular to the local EHR environment.

Outside of these formatting issues, we found significant variation within clinical documents even at the section level. For instance, within a subset of four or five document types within our corpus, over 25,000 lexically unique section headers were encountered (even after normalization for capitalization). Furthermore, there were additional errors owing to the lack of spell-checking [21] or language/grammar mistakes. Dictation from voice transcription may also result in mistakes because of a misinterpretation of word meanings/intentions between the clinician's intended meaning and the interpretation by the transcriptionist. Additionally, clinicians often used sentence fragments instead of fully structured sentences for efficient communication. This custom may hinder automatic WSD research because the extracted features miss valuable information.

2) Challenge 2: Shortage of resources

Another significant issue involved the currently limited resources and knowledge for WSD research in the clinical domain. For example, there are only a few available clinical acronym and abbreviation datasets (e.g., datasets by Xu et al. [15] or Mayo Clinic set [14]). Currently, there are no comprehensive clinical sense inventories for large numbers of acronyms available. Furthermore, it is well known there is a bottleneck problem in knowledge acquisition when collecting data or aggregating valuable information. In other words, the significant time, cost, and effort of experts are indispensable to obtain useful knowledge.

3) Challenge 3: Privacy issues

Finally, maintaining privacy and security while allowing greater access remains a significant challenge for researchers wishing to utilize clinical notes. Patient confidentiality policies with the Health Insurance Portability and Accountability Act (HIPAA) are strict. Large corpora of clinical text have not been traditionally easily available to NLP researchers. Rare exceptions include several notable efforts, including the i2b2 challenges [22] and the University of Pittsburgh Medical Center (UPMC) de-identified clinical notes repository [23]. Even the i2b2 and the UPMC datasets are not free of restrictions and necessary regulatory approvals. Furthermore, researchers with potential access to clinical documents containing protected health information at their own institutions encounter significant political and regulatory issues in gaining access to these documents or sharing these documents [3,16]. Consequently, relatively little research has been done on acronyms and abbreviations in clinical notes compared to biomedical literature documents. At our institution, we were able to obtain Institutional Review Board (IRB) approval for our research and to work directly with our clinical partners at the hospital to perform this research.

2. Starting from Simple Approaches

To overcome some of these difficulties, our approach has been to utilize the available resources, tools, and knowledge of other interdisciplinary fields. Overlapping concepts of the biomedical fields and domain-independent approaches to analyzing English language usage from the linguistic or statistical fields are methods that are reasonable to start with that might work directly or with some adaptation to the clinical domain. For feature selection, we expected that the utilization of advanced/elaborate tools, techniques, and knowledge especially from the biomedical fields would be beneficial for automatic WSD research in clinical documents.

These tools, however, have not been optimized for use within the clinical domain. For example, in contrast to biomedical literature discourse, as previously mentioned, clinical documents have short forms that are rarely associated with long forms. Moreover, biomedical tools utilize sense inventories primarily derived from biomedical literature, often ignoring or missing important clinical senses. Liu et al. [24] reported that UMLS covers only 66% of acronyms and abbreviations with less than 6 characters in the clinical domain. Many times, at least one of the senses encountered in our corpus was not contained within available references, such as the UMLS [4,5,16,24] or biomedical sense inventories, such as Adam [25]. Also, biomedical tools and typical linguistic tools (i.e., tokenizers or POS taggers) may fail with clinical text, where statements are often fragmented, sometimes without proper sentence boundaries.

The balance between utilizing existing tools, some of which have limited options for modification, versus development of customizable tools, remains difficult. We have attempted to use existing established tools where available, and retrain these tools where possible with clinical text. We are also trying to understand where these resources succeed and fail in order to optimize the previous work of others and reuse these resources. Moreover, we are focusing our efforts in areas where highly-specific tools for medical text are needed. In these attempts, it seems that clinically-oriented approaches would be helpful. Adopting clinical cognitive flows from medical specialties, position in discourse, and section information may be helpful.

Using simple features with simple ML algorithms can be a reasonable starting point. As a starting point, we chose to use the BoW approach to feature extraction coupled with the SVM algorithm. BoW was found to produce a simple but effective feature set. In our research, BoW often demonstrated competitive results over other isolated features or combinations of feature sets when various ML algorithms were used. Table 1 depicts the clinical sense distribution of seven abbreviations from our set. Table 2 shows the performance of different feature sets with naïve Bayes, SVMs, and decision tree algorithms for each of the seven abbreviations. Moreover, the combination of all possible feature sets (BoW, CUI, semantic type, and section) deteriorates the performance of ML because of duplicate and unnecessary data diluting critical information for ML algorithms in our experience (not pictured). As such, combining heterogeneous information with the prevention of overfitting is necessary.

Table 1. The sense distribution of abbreviations

Abbreviation	No. of instances	Coverage	Sense
BK (2 senses)	343	0.686	BK (virus)
	157	0.314	Below knee
C3 (5 senses)	245	0.490	Cervical 3
	235	0.470	Component 3
	20	0.040	3 more other senses
C4 (6 senses)	254	0.508	Cervical 4
	229	0.458	Component 4
	17	0.034	4 more other senses
CVA (2 senses)	278	0.556	Cerebrovascular accident
	222	0.444	Costovertebral angle
ET (8 senses)	290	0.580	Enterostomal therapy
	198	0.396	Endotracheal
	12	0.024	6 more other senses
GC (4 senses)	310	0.620	*Gonococcus*
	184	0.368	Genetic counselor
	6	0.012	2 more other senses
IA (6 senses)	296	0.592	IA
	175	0.350	Intra-arterial
	29	0.058	4 more other senses

Number of instance is the acronym having the given sense; Coverage is the percentage of the particular abbreviations.

http://dx.doi.org/10.4258/hir.2015.21.1.35

Table 2. Precision, recall, and F-measure for abbreviations

Abbreviation	Features	NB			SVM			DT		
		Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
BK	Majority	0.471	0.686	0.558	0.471	0.686	0.558	0.471	0.686	0.558
	BoW	0.984	0.984	0.984	0.980	0.980	0.980	0.960	0.960	0.960
	CUI	0.910	0.910	0.910	0.904	0.900	0.896	0.906	0.894	0.887
	Semantic type	0.841	0.834	0.836	0.849	0.850	0.850	0.825	0.824	0.824
	Section	0.811	0.816	0.809	0.811	0.816	0.809	0.824	0.826	0.817
C3	Majority	0.240	0.490	0.322	0.240	0.490	0.322	0.240	0.490	0.322
	BoW	0.948	0.954	0.948	0.939	0.946	0.938	0.842	0.850	0.846
	CUI	0.775	0.726	0.724	0.817	0.814	0.806	0.739	0.716	0.691
	Semantic type	0.778	0.762	0.760	0.785	0.778	0.768	0.743	0.750	0.743
	Section	0.726	0.738	0.731	0.708	0.732	0.717	0.710	0.734	0.719
	BoW + CUI	0.943	0.942	0.935	0.944	0.946	0.937	0.843	0.850	0.846
C4	Majority	0.258	0.508	0.342	0.258	0.508	0.342	0.258	0.508	0.342
	BoW	0.952	0.954	0.952	0.949	0.952	0.947	0.867	0.880	0.872
	CUI	0.796	0.760	0.748	0.844	0.842	0.836	0.772	0.756	0.735
	Semantic type	0.765	0.712	0.733	0.723	0.746	0.733	0.708	0.730	0.719
	Section	0.726	0.706	0.713	0.733	0.742	0.734	0.730	0.738	0.730
	BoW + CUI	0.957	0.958	0.956	0.949	0.952	0.946	0.863	0.876	0.868
CVA	Majority	0.309	0.556	0.397	0.309	0.556	0.397	0.309	0.556	0.397
	BoW	0.974	0.974	0.974	0.970	0.970	0.970	0.974	0.974	0.974
	CUI	0.961	0.960	0.960	0.947	0.946	0.946	0.934	0.930	0.929
	Semantic type	0.914	0.914	0.914	0.910	0.910	0.910	0.928	0.928	0.928
	Section	0.950	0.950	0.950	0.952	0.952	0.952	0.948	0.948	0.948
	BoW + CUI	0.972	0.972	0.972	0.976	0.976	0.976	0.974	0.974	0.974
ET	Majority	0.336	0.580	0.426	0.336	0.580	0.426	0.336	0.580	0.426
	BoW	0.941	0.962	0.951	0.935	0.958	0.946	0.939	0.950	0.942
	CUI	0.918	0.924	0.918	0.931	0.940	0.932	0.918	0.924	0.918
	Semantic type	0.899	0.894	0.896	0.901	0.910	0.905	0.873	0.884	0.878
	Section	0.795	0.788	0.792	0.760	0.780	0.769	0.765	0.782	0.769
	BoW + CUI	0.946	0.952	0.946	0.951	0.962	0.953	0.926	0.948	0.937
GC	Majority	0.384	0.620	0.475	0.384	0.620	0.475	0.384	0.620	0.475
	BoW	0.984	0.988	0.985	0.986	0.990	0.987	0.977	0.980	0.977
	CUI	0.863	0.798	0.799	0.879	0.834	0.835	0.825	0.706	0.700
	Semantic type	0.890	0.874	0.877	0.892	0.892	0.888	0.875	0.872	0.868
	Section	0.947	0.956	0.951	0.956	0.966	0.961	0.958	0.970	0.964
	BoW + CUI	0.986	0.990	0.988	0.982	0.986	0.982	0.977	0.980	0.977
IA	Majority	0.350	0.592	0.440	0.350	0.592	0.440	0.350	0.592	0.440
	BoW	0.941	0.944	0.941	0.959	0.962	0.958	0.904	0.914	0.905
	CUI	0.797	0.758	0.762	0.808	0.820	0.803	0.726	0.758	0.728
	Semantic type	0.773	0.728	0.747	0.778	0.778	0.768	0.719	0.734	0.721
	Section	0.638	0.602	0.578	0.650	0.652	0.587	0.675	0.662	0.590
	BoW + CUI	0.930	0.932	0.929	0.958	0.960	0.954	0.910	0.918	0.909

NB: naïve Bayes, SVM: support vector machine, DT: decision tree, BoW: bag of words, CUI: concept unique identifier.

3. Considerations in Assessing the Complexity of each Acronym and Abbreviation WSD Task

Even if we only use the simple BoW approach, it can be difficult to detect patterns when acronyms and abbreviations have different or skewed sense distributions. In our research to date looking at a large sample of acronyms and abbreviations within clinical documents, approximately half of these acronyms have only a single meaning (long form or sense) contained within a random moderate-sized sample of instances of each acronym in question (500 occurrences). After excluding acronyms with a single sense or a locally specific (specific to a particular institution) sense, we have found that biomedical sense inventories have a significant level of redundancy/synonymy between different long form expressions. This has required additional steps to reduce the redundancy of sense in sense inventories that needed to be taken prior to WSD [26].

After taking into account these factors, we observed performance patterns when grouping acronyms and abbreviations based on the majority sense rate. For instance, we found similar performance when the majority sense rate is relatively balanced because we are able to gather enough information about minority senses due to the availability of samples for each of the senses. On the other hand, gaining sufficient performance improvements can be difficult if distribution is skewed because the small number of samples for each of the minority samples may give insufficient information to help with proper disambiguation.

Another convenient consideration for researchers is an understanding of the relative degree of 'separateness' between senses. 'Well-separated' typically implies substantial semantic differentiation among senses [27]. However, 'well-separated' senses within a clinical text may also imply different uses within various sections of a single clinical note (e.g., different relative note location or different section). In principle, well-separated senses should then yield a higher accuracy for classification or clustering [4,28,29]. For instance, 'CVA' has two well-separated senses, 'cerebrovascular accident' and 'costovertebral angle'. In our experience, most of the feature sets for 'CVA' perform high accuracies, over 95%, when the SVM algorithm is used with only a few dozen samples during the training phases. Therefore, depending on the degree of separation, the necessary sample size for a training phase to achieve good performance might be estimated [4,30]. We have found to date a simple trend that acronyms or abbreviations having a single highly prevalent sense need more training samples than acronyms or abbreviations having evenly distributed senses to achieve significant performance improvements.

IV. Discussion

The proliferative use of acronyms and abbreviations in the clinical domain makes automatic sense disambiguation of acronyms and abbreviations for medical NLP systems an important ongoing challenge and area of research. For this pilot research, we studied WSD tasks for a few acronyms and abbreviations from clinical notes. From this, we have learned the following lessons: 1) practical difficulties with using clinical text that must be solved including language, structure and formatting issues, as well as lack of resources for clinical text and privacy issues; 2) starting from simple approaches, such as single features using well-known ML algorithms or using of well-separated senses is a sensible initial approach; and 3) to understand the performance of ML algorithms better, one should consider the distribution of senses of an acronym or abbreviation as well as the degree of separateness between senses and of usage between different long forms of an acronym. After these simple approaches, we need to customize tools and knowledge in order to harmonize clinical resources or to develop new tools from clinical fields.

According to several literature reviews examining biomedical and clinical documents, an optimal feature or set of features that will adequately address the disambiguation problem for biomedical or general English acronyms and abbreviation has not been found [9,16,30]. Due to these factors, accomplishing representative and optimal feature selection, a key step for classification or clustering, is an area of open research. Even if there are particular advantages and disadvantages of individual features within the clinical domain, there is no absolute superior feature identified for this task to date. Further study is needed with careful consideration of overfitting in clinical acronym and abbreviation WSD.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by the American Surgical Association Foundation Fellowship, the University of Minnesota Institute for Health Informatics Seed Grant, and by the National Library of Medicine (#R01 LM009623-01). We would like to thank Fairview Health Services for support of this research. The authors also thank Serguei Pakhomov PhD for insightful comments.

References

1. Pakhomov S. Semi-supervised Maximum Entropy based approach to acronym and abbreviation normalization in medical texts. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL2002); 2002 Jul 6-12; Philadelphia, PA. p. 160-7.
2. Stetson PD, Johnson SB, Scotch M, Hripcsak G. The sublanguage of cross-coverage. Proc AMIA Symp 2002; 2002:742-6.
3. Pakhomov S, Pedersen T, Chute CG. Abbreviation and acronym disambiguation in clinical discourse. AMIA Annu Symp Proc 2005:589-93.
4. Xu H, Markatou M, Dimova R, Liu H, Friedman C. Machine learning and word sense disambiguation in the biomedical domain: design and evaluation issues. BMC Bioinformatics 2006 5;7:334.
5. Xu H, Stetson PD, Friedman C. A study of abbreviations in clinical notes. AMIA Annu Symp Proc 2007;2007: 821-5.
6. Kuhn IF. Abbreviations and acronyms in healthcare: when shorter isn't sweeter. Pediatr Nurs 2007;33:392-8.
7. Walsh KE, Gurwitz JH. Medical abbreviations: writing little and communicating less. Arch Dis Child 2008;93(10): 816-7.
8. Hunt DR, Verzier N, Abend SL, Lyder C, Jaser LJ, Safer N, et al. Fundamentals of medicare patient safety surveillance: intent, relevance, and transparency. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. Advances in patient safety: from research to implementation (Volume 2: Concepts and Methodology). Rockville (MD): Agency for Healthcare Research and Quality; 2005.
9. Fan JW, Friedman C. Word sense disambiguation via semantic type classification. AMIA Annu Symp Proc 2008;2008:177-81.
10. Friedman C, Liu H, Shagina L, Johnson S, Hripcsak G. Evaluating the UMLS as a source of lexical knowledge for medical language processing. Proc AMIA Symp 2001;2001:189-93.
11. Schuemie MJ, Kors JA, Mons B. Word sense disambiguation in the biomedical domain: an overview. J Comput Biol 2005;12:554-65.
12. Kaplan A. An experimental study of ambiguity and context. Mech Transl 1950;2(2):39-46.
13. Choueka Y, Lusignan S. Disambiguation by short contexts. Comput Hum 1985;19(3):147-57.
14. Joshi M, Pakhomov S, Pedersen T, Chute CG. A comparative study of supervised learning as applied to acronym expansion in clinical reports. AMIA Annu Symp Proc 2006;2006:399-403.
15. Xu H, Stetson PD, Friedman C. Methods for building sense inventories of abbreviations in clinical notes. J Am Med Inform Assoc 2009;16(1):103-8.
16. Savova GK, Coden AR, Sominsky IL, Johnson R, Ogren PV, de Groen PC, et al. Word sense disambiguation across two domains: biomedical literature and clinical notes. J Biomed Inform 2008;41(6):1088-100.
17. Manning CD, Schutze H. Foundations of statistical natural language processing. Cambridge (MA): MIT Press; 1999.
18. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001;2001:17-21.
19. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Stud Health Technol Inform 2001;84(Pt 1):216-20.
20. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD Explor 2009;11(1):10-8.
21. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform 2008;2008:128-44.
22. NLP research data sets [Internet]. Boston (MA): i2b2; c2014 [cited at 2015 Jan 5]. Available from: https://www.i2b2.org/NLP/DataSets/Main.php.
23. University of Pittsburgh NLP Repository [Internet]. Pittsburgh (PA): Department of Biomedical Informatics, University of Pittsburgh; c2014 [cited at 2015 Jan 5]. Available from: http://www.dbmi.pitt.edu/nlpfront.
24. Liu H, Lussier YA, Friedman C. A study of abbreviations in the UMLS. Proc AMIA Symp 2001;2001:393-7.
25. Zhou W, Torvik VI, Smalheiser NR. ADAM: another database of abbreviations in MEDLINE. Bioinformatics 2006;22(22):2813-8.
26. Melton GB, Moon S, McInnes B, Pakhomov S. Automated identification of synonyms in biomedical acronym sense inventories. Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents; 2010 Jun 5; Los Angeles, CA. p. 46-52.
27. Resnik P, Yarowsky D. Distinguishing systems and distinguishing senses: new evaluation methods for word sense disambiguation. Nat Lang Eng 1999;5(2):113-33.
28. Leroy G, Rindflesch TC. Using symbolic knowledge in the UMLS to disambiguate words in small datasets with a naïve Bayes classifier. Stud Health Technol Inform

2004;107(Pt 1):381-5.
29. Stevenson M, Guo Y, Amri AA, Gaizauskas R. Disambiguation of biomedical abbreviations. Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing (BioNLP); 2009 Jun 4-5; Boulder, CO. p. 71-9.
30. Liu H, Teller V, Friedman C. A multi-aspect comparison study of supervised word sense disambiguation. J Am Med Inform Assoc 2004;11(4):320-31.