ISSN 1870-4069

A Supervised Learning for Assigning Categories to Medical Concepts and Contexts

Anupam Mondal[1], Erik Cambria[2],
Dipankar Das[1], Sivaji Bandyopadhyay[1]

[1] Jadavpur University, Kolkata,
Department of Computer Science and Engineering,
India

[2] Nanyang Technological University,
School of Computer Science and Engineering,
Singapore

anupam@sentic.net, dipankar.dipnil2005@gmail.com,
sbandyopadhyay@cse.jdvu.ac.in, cambria@ntu.edu.sg

Abstract. Category assignment of medical concepts is presented as an essential task in information extraction from a large number of available medical corpora in the form of unstructured and semi-structured. Besides, the category of medical concepts helps to identify the conceptual knowledge from the corpora, especially in healthcare. So, in the present task, we are motivated to design a category assignment system for the medical concepts and contexts which assists in understanding the subjective information of the medical corpus. The proposed system is able to identify five different types of categories for medical concepts (e.g., *diseases, drugs, symptoms, human_anatomy,* and *miscellaneous medical terms (MMT)*) and eleven types of context categories (e.g., *disease-symptom, disease-drug,* and *disease-MMT*). In order to design the system, we have distributed the task into three subtasks such as medical concept identification, concept category assignment, and category assignment for the context. Moreover, we have employed a domain-specific lexicon namely WordNet of Medical Event (WME 3.0) to design all three subtasks. The lexicon assists in recognizing the medical concepts and assigning the categories of these concepts using its provided various features as affinity score, gravity score, polarity scores, similar sentiment words, and sentiment. Thereafter, we have used two well-known supervised classifiers to build and validate the proposed both of the category assignment systems.

Keywords: Medical concepts, medical text categories, text classifier.

1 Introduction

In healthcare, category assignment of medical concepts and contexts are challenging due to the lack of involvement of domain-experts as well as unavailability of domain-specific lexicons and ontologies [27, 37].

In order to address these challenges, the researchers have developed several information extraction systems as Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT), Unified Medical Language System (UMLS), and GENIA [4]. The systems help to recognize the medical concepts and their various conceptual features from unstructured corpora which assists in presenting structured corpus from unstructured corpora [18, 9].

On the other side, the researchers have designed various statistical and ontology-based approaches to identify the information automatically from the corpus [20]. The automated systems help the experts' like doctors and medical practitioners and non-experts as patients to describe the information related to medical concepts and contexts.

The concepts refer the key medical terms where the contexts represent the sentences of the corpus. For example, the medical concept "congenital_heart_disease" and medical context "A congenital heart defect (CHD) is a problem in the structure of the heart that is present at birth." respectively.

In the present paper, we are motivated to design an automated information extraction system namely categorization of medical concepts and contexts with the help of two well-known machine learning classifiers viz. Naïve Bayes and Logistic Regression [11]. Besides, we have prepared an experimental dataset and extracted features for medical concepts such as statistical, linguistic, and syntactical to build and validate both of the systems [36, 3].

We have also observed that a domain-specific lexicon is essential to extract the subjective and conceptual features of the medical concepts. Hence, we have employed previously developed a domain-specific lexicon viz. WordNet of Medical Event (WME 3.0) to design the proposed systems. The lexicon assists in recognizing the medical concepts in a context and extracting the features such as affinity score, category, gravity score, polarity score, Parts-Of-Speech (POS), sentiments, and similar sentiment words (SSW) for the medical concepts [26, 25].

We have also noticed the following difficulties exists to design the proposed categorization system. ***A.*** how to recognize the medical concepts (e.g., *breathing_problem*) and isolate the general concepts (e.g., organization) from the corpus? ***B.*** how to decide the set of medical categories and assign them to the medical concepts? This challenge is also presented as a classification task of medical concepts [28] viz.

A chronic cough is present either *symptom* or *disease* category. ***C.*** how to assign the set of categories for medical contexts and assign them? ***E.*** finally, how to prepare an experimental dataset to design and validate the proposed system? In order to address the mentioned challenges, we have used WME 3.0 lexicon which helps to recognize the medical concepts and isolate from non-medical concepts.

Thereafter, we have proposed five different types of categories for medical concepts such as *disease, symptom, drug, human_anatomy,* and *miscellaneous terms (MMT)* as unrecognized categories, suggested by medical practitioners after observing the experimental dataset. Besides, these categories assist in presenting eleven different pair-based categories for the medical context such as *disease-symptom*, *disease-drug*, *disease-human_anatomy*, *disease-MMT*, *symptom-drug*, *symptom-human_anatomy*,

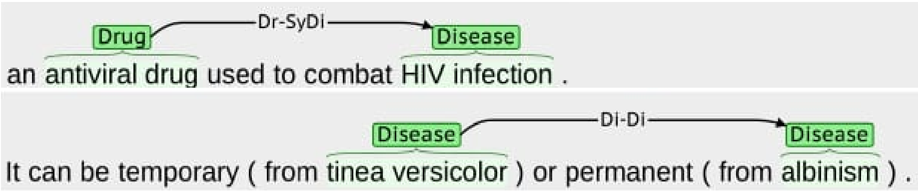

Fig. 1. A sample output of the proposed categorization system.

symptom-MMT, *human_anatomy-MMT*, *drug-human_anatomy*, *drug-MMT*, and *MMT-MMT*. Thereafter, we have employed Naïve Bayes and Logistic Regression classifiers along with various extracted features to develop the proposed categorization system.

The systems have provided F-score 0.81 and 0.86 for assigning the categories of medical concepts and contexts individually. Figure 1 shows a sample output of the proposed system.

We have observed that the categories of medical concepts and contexts facilitate to identify the subjective and conceptual information from the corpus. The system also helps to build various applications like medical concept network, annotation system, and recommendation system in healthcare services [22, 8, 9].

The rest of the paper is as follows. Section 2 presents the background of this work. Section 3 describes the preparation process of the experimental dataset and a brief overview WME lexicon. Section 4 discusses how we have developed both of the categorization systems such as concepts and contexts. Finally, the concluding remarks appear in Section 5.

2 Background

Information extraction from the biomedical textual corpus is demanding due to the lack of domain experts involvement and unavailability of structured corpora [21, 6]. In order to produce the structured corpus from the largely available semi-structured and unstructured corpora, the researchers have developed several ontologies and lexicons, which are also used for building automated information extraction system.

Few of the ontologies and lexicons are UMLS (Unified Medical Language System) and SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms), MEN (Medical WordNet), and WME (WordNet of Medical Event) [33, 15, 25]. Primarily, these resources are looking for recognizing the conceptual as well as knowledge-based information from corpora [13, 5].

In the current research, we have focused on designing an automated categorization system using machine learning approaches in the presence of a domain-specific lexicons. So, in the following, we have discussed how the lexicons are taking a crucial role to develop the category assignment system for the medical concepts in healthcare.

Medical WordNet (MEN) lexicon built with two different sub-networks viz. Medical FactNet (MFN) and Medical BeliefNet (MBN), to evaluate consumer health reports [33].

The formal architecture of Princeton WordNet has been used to develop this lexicon [15]. On the other side, Mondal et al., [26, 25] have designed a domain-specific lexicon like WordNet of Medical Events (WME 3.0) to identify medical concepts and their various features. The features are affinity score, categories, gloss, gravity score, Parts-Of-Speech, polarity score, sentiment, and similar sentiment words. In this work, we have employed this lexicon to prepare and build the automated category assignment system for medical concepts and contexts.

Besides, we have noticed that the fundamental and broad classes such as category identification are still challenging due to insufficient number of domain experts involvement [23]. We have observed few of the category assignment research as follows: Yao, et al., [38] have extracted category-based relations as *cures, prevents,* and *side_effects*, which describe the distinctive nature from the biomedical text (medical papers) [2, 17].

Franzen et al., [16] have annotated Yapex corpus with 200 medical abstracts to assign the category such as *proteins*. Primarily, the ontology looks for extracting *protein-protein* interaction and *disease-treatment* relations from corpora under a BioText project [1, 30, 31, 19]. Eklund [12] developed an annotation system to extract the relations as *diseases* for *treatments* from the scientific medical corpus.

The literature survey helps to motivate to design the category assignment system for medical concepts and contexts both. Hence, we have used WME 3.0 lexicon and two well-known classifiers such as Naïve Bayes and Logistic Regression to prepare the experimental dataset, design, and validate the system. The concept categories show the subjective information for each of the medical concepts whereas context categories present the conceptual relation between each pair of medical concepts in a context.

3 Resource and Dataset

The present section describes how we employed WordNet of Medical Event (WME 3.0) lexicon to prepare the experimental dataset for the proposed categorization system.

*WordNet of Medical Event Lexicon:*A domain-specific lexicon is demanding to identify the conceptual information such as category or sentiment from the medical corpus [7]. So, we have borrowed the knowledge from the current version of WordNet of Medical Event lexicon namely WME 3.0 [26, 25, 24].

WME 3.0 lexicon was developed with an additional category feature of 10186 number of medical concepts along with previous two versions such as WME 1.0 and WME 2.0. WME 1.0 is able to identify the gloss (descriptive explanation), Parts-Of-Speech (POS), polarity score, and sentiment for 1654 number of medical concepts where WME 2.0 was added affinity score, gravity score, and Similar Sentiment Words (SSW) features for 6415 number of medical concepts.

The conventional WordNet[1], a preprocessed medical dictionary, and SenticNet[2] resources were applied to prepare all three versions of WME [10, 35]. In the following, we have discussed how the affinity score, gravity score, and category features help to design the proposed system.

[1] https://wordnet.princeton.edu/

[2] http://sentic.net/

```
<?xml version="3.0" encoding="UTF-8"?>
- <Medical Concepts>
 - <Concept>
     ........
   </Concept>
 - <Concept>
     <Title>amnesia</Title>
   - <Properties>
       <POS>noun</POS>
       <Category>disease</Category>
       <Gloss>Loss of memory sometimes including the memory of
        personal identity due to brain injury, shock, fatigue, repression,
        or illness or sometimes induced by anesthesia.</Gloss>
       - <SSW and Affinity score>
              blackout (0.674)
              memory loss (0.534)
              stupor (0.429)
              fugue (0.345)
         </SSW and Affinity score>
       <Polarity score> - 0.375</Polarity score>
       <Gravity score>0.170</Gravity score>
       <Sentiment>negative</Sentiment>
     </Properties>
   </Concept>
 - <Concept>
     ........
   </Concept>
 </Medical Concepts>
```


Fig. 2. The current version of WordNet of Medical Event (WME 3.0) lexicon. The features considered in WME 3.0, affinity score, category, gloss, gravity score, Parts-Of-Speech (POS), polarity score, sentiment, and Similar Sentiment Words (SSW) are listed for an example medical concept ***amnesia***.

1. Affinity score indicates what extend two concepts are close to each other by measuring the number of common sentiment words (SSW) appeared for a pair of concepts within the range of 0 to 1.
2. Gravity score identifies sentiment-oriented relevance between medical concepts and their various glosses (descriptive explanations) and ranges from -1 to 1. While -1 suggests no relation and 1 indicates strong relations either positive or negative, which helps to identify a proper gloss for concepts.
3. Besides, the assigned categories such as *diseases, drugs, treatments, human_anatomy,* and *MMT* assist in extracting the subjective information of the concepts.

In the present paper, WME 3.0 lexicon has been used as a baseline to build a category assignment system. Figure 2 shows a sample output of WME 3.0 lexicon.

Experimental Dataset: A label dataset is essential to build and validate any information extraction system with machine learning classifiers. Hence, we have acquired and prepared an experimental dataset from two different resources such as one SemEval 2015 Task-6[3] and two MedicineNet[4].

[3] http://alt.qcri.org/semeval2015/task6/

[4] http://www.medicinenet.com/script/main/hp.asp

Table 1. A comparative statistics between two resources namely SemEval 2015 Task-6 and MedicineNet which are used to prepare the experiment dataset.

	SemEval 2015 Task-6	MedicineNet
Total number of contexts (both medical and non-medical)	4023	3329
Unique number of Medical contexts	2481	2583
Total number of concepts (only medical)	14665	16809
Unique number of Medical concepts	4485	4847
Average number of Medical concepts per contexts (respect to unique)	1.81	1.88
No. of mentioned categories	3	4

Table 2. A detail statistics of the experimental dataset.

Statistics of the labelled Experimental dataset				
		Experimental	Training (40%)	Test (60%)
Context Categories	*Disease-Symptom*	938	375	563
	Disease-Drug	713	285	428
	Disease-Human_anatomy	315	126	189
	Disease-MMT	535	214	321
	Symptom-Drug	108	43	65
	Symptom-Human_anatomy	78	31	47
	Symptom-MMT	135	54	81
	Human_anatomy-MMT	105	42	63
	Drug-Human_anatomy	285	114	171
	Drug-MMT	117	47	70
	MMT-MMT	1735	694	1041
Concept Categories	*Diseases*	2246	898	1348
	Symptoms	1204	482	722
	Drugs	2108	843	1265
	Human_anatomy	428	171	257
	MMT	3346	1338	2008

The experimental dataset contains 5064 and 9332 number of unique medical contexts and concepts individually after combining both of the resources. Table 1 presents a statistical comparison between these resources.

Besides, a group of medical practitioners helps to label both of the experimental datasets in the presence of WME 3.0 lexicon. Thereafter, we have split the datasets into two parts such as training (40%) and test (60%) dataset. The training dataset uses for building the systems where test dataset applies for validating them. Table 2 shows the distributions of both of the experimental datasets in details.

In the following sections, we have discussed how the prepared experimental dataset assists in building the categorization system for medical concepts and contexts.

4 Category Assignment Systems

The conceptual classes of medical concepts present the categories which help to identify the conceptual information from the corpus [32].

Both of the categories such as concepts and contexts of medical concepts assist in representing structured corpus from a large number of the unstructured and semi-structured corpora produced by the medical practitioners. Besides, these categories help to design various decision-making systems in healthcare services as annotation and recommendation etc [34, 29]. Hence, in the present work, we have recognized five different types of categories of medical concepts and their related eleven types of contexts categories.

The categories of medical concepts are *disease, drug, human_anatomy, symptom* and *MMT* where the categories of contexts are *disease-symptom*, *disease-drug*, *disease-human_anatomy*, *disease-MMT*, *symptom-drug*, *symptom-human_anatomy*, *symptom-MMT*, *human_anatomy-MMT*, *drug-human_anatomy*, *drug-MMT*, and *MMT-MMT*. We have proposed categories of the context as a pair-based due to the short length in the experimental dataset.

Thereafter, to assign the mentioned categories for both concepts as well as contexts, we have proposed two different categorization systems using machine learning classifiers individually. First categorization system is able to identify the subjective information from the corpus whereas the second system is able to extract the overall conceptual information from the corpus. In the following subsections discuss, how we have developed and validated both of the categorization systems in details.

4.1 Concept Categorization System

In order to design the proposed category assignment system for the medical concept, we have extracted various linguistic as well as knowledge-based features with two well-known supervised classifiers viz. Naïve Bayes and Logistic Regression. The features are POS, uni-gram, bi-gram, tri-gram, sentiment, and Similar Sentiment Words (SSW). To extract these features, we have employed WME 3.0 lexicon along with conventional WordNet, SentiWordNet, and SenticNet resources [14, 10].

Initially, we have extracted the mentioned features for the training dataset as mentioned in Section 3 and process it through two different classifiers individually. Thereafter, we have combined both of the classification modules to build the final category assignment system for medical concepts. The following steps discussed the system development in details.

- **Step-1:** Extract various features for the medical concepts of the training dataset using the mentioned resources (e.g., WME 3.0 and WordNet) and prepare a feature vector, F_V = [(1,0,1,1,1,1), (1,1,0,1,0,1), (1,1,0,1,0,0), ..., (1,1,1,1,0,1)].
- **Step-2:** Assign five different category label C_L = [1, 2, 2, ..., 4] to the feature vector F_V. The category labels are 1, 2, 3, 4, and 5 for *diseases, symptoms, drugs, human_anatomy*, and *MMT* individually.
- **Step-3:** The feature vector F_V and its corresponding label C_L have applied through two different classifiers such as Naïve-Bayes and Logistic regression and build two approaches as M_{NB} and M_{LR} respectively.
- **Step-4:** Thereafter, we have combined these approaches M_{NB} and M_{LR} with the union operator and built another approach ($M_{Combined}$) under the proposed module.

Table 3. An ablation study (as F-score) for assigning categories of medical concepts using three approaches viz. Naïve Bayes, logistic regression, and combined.

Features	M_{NB}			M_{LR}			$M_{Combined}$		
	Precision	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
n-grams	0.79	0.91	0.85	0.80	0.89	0.84	0.81	0.83	0.85
n-grams + POS	0.82	0.89	0.85	0.83	0.87	0.85	0.84	0.88	0.86
Sentiment + SSW	0.75	0.83	0.79	0.77	0.85	0.81	0.78	0.84	0.81
POS + Sentiment + SSW	0.83	0.89	0.86	0.84	0.90	0.87	0.85	0.91	0.88
n-grams + Sentiment + SSW	0.84	0.88	0.86	0.83	0.89	0.86	0.86	0.91	0.88

Table 4. A comparative result analysis to assign categories of medical concepts using Naïve Bayes, logistic regression, and their combined approach.

	M_{NB}			M_{LR}			$M_{Combined}$		
	Precision	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
diseases	0.85	0.97	0.91	0.85	0.97	0.91	0.86	0.98	0.92
symptoms	0.87	0.78	0.82	0.88	0.79	0.83	0.91	0.79	0.85
drugs	0.91	0.87	0.89	0.90	0.88	0.89	0.92	0.89	0.90
human_anatomy	0.67	0.90	0.77	0.70	0.88	0.78	0.67	0.91	0.77
MMT	0.79	0.84	0.81	0.81	0.84	0.82	0.86	0.81	0.83

For example, the proposed module is able to identify the medical concepts ***pneumonia*** and ***infection*** and their category as *disease* and *symptom* from the medical context "***Pneumonia*** *is frequently but not always due to* ***infection***".

4.2 Validation of Concept Categorization System

Thereafter, to validate the output of the proposed system, we have used the labeled test dataset as mentioned in Section 3. We have processed the test dataset with three modules viz. M_{NB}, M_{LR}, and $M_{Combined}$ consequently to calculate the accuracy of the final categorization system for medical concepts. Table 3 shows the importance of the applied features through an ablation study for the proposed system. We have also conducted a comparative result analysis for all three modules by calculating Precision, Recall, and F-score as shown in Table 4.

Finally, we have compared the category output of the proposed system along with WME 3.0 lexicon to prove the accuracy of the proposed system as shown in Table 5. We have also noticed that the coverage of the extracted categories of medical concepts is same with WME 3.0 lexicon due to use of the similar type of resources.

4.3 Context Categorization System

In addition, the category assignment system for the medical contexts is essential to determine the overall conceptual information for the corpus. So, we have proposed another categorization system on the top of the categorization system of medical concepts. The context categorization system is able to identify eleven types of pair-based categories for the medical corpus. Primarily these categories are derived after observing our experimental dataset by a group of medical practitioners.

Table 5. A statistical comparison of the coverage between the extracted categories from test dataset of the proposed system and WME 3.0 lexicon.

Distributions	Test dataset			WME 3.0
	Extracted	**Unique**	**Coverage (%)**	**Unique**
Disease	1321	1248	100	3641
Symptom	570	534	100	802
Drug	1126	1079	100	4196
Human_anatomy	234	206	100	227
MMT	1627	1185	85.07	1320

Table 6. A comparative result analysis to assign the categories of medical contexts using the proposed system.

	Test Dataset	Extracted	Precision	Recall	F-score
Disease-Symptom	563	545	0.95	0.98	0.96
Disease-Drug	428	397	0.89	0.96	0.92
Disease-Human_anatomy	189	165	0.80	0.92	0.86
Disease-MMT	321	278	0.82	0.95	0.88
Symptom-Drug	65	46	0.58	0.83	0.68
Symptom-Human_anatomy	47	32	0.62	0.91	0.74
Symptom-MMT	81	66	0.65	0.80	0.72
Human_anatomy-MMT	63	38	0.49	0.81	0.61
Drug-Human_anatomy	171	152	0.84	0.94	0.89
Drug-MMT	70	43	0.53	0.86	0.66
MMT-MMT	1041	971	0.77	0.83	0.80

Thereafter, to design the proposed categorization system for medical contexts, we have applied the following algorithm on training dataset in the presence of concept categorization system.

Step-1: Initially, we have assigned the categories of medical concepts in a context using concept categorization system. Each medical concept and its category presented as $Medi_C$ in a context.

Step-2: Consider the consecutive medical concepts and their category from the context, which is presented as Partial Context Category (PCC).

Step-2.1: If the consecutive pair of concept categories are same then PCC is:

$$PCC = Medi_{C1} \cap Medi_{C2}. \tag{1}$$

Step-2.2: Else PCC is:

$$PCC = Medi_{C1} \cup Medi_{C2}, \tag{2}$$

where $Medi_{C1}$ and $Medi_{C2}$ indicate two consecutive medical concepts and their categories.

Step-3: The extracted Partial Context Category (PCC) helps to assign the Final Context Category (FCC) as:

$$FCC = PCC_1 \cap PCC_2, \tag{3}$$

where PCC_1 and PCC_2 refer the partial context category in a context.

For example, the proposed system is able to assign *human_anatomy-MMT* category for the following medical context after identifying the medical concepts and their corresponding categories such as "***passage***" as *MMT*, "***air***" as *MMT*, "***supply***" as *MMT*, "***oxygen***" as *MMT*, "***lungs***" as *human_anatomy*, and "***body***" as *human_anatomy*.

"*The passage of air into and out of the lungs to supply the body with oxygen*".

4.4 Validation of Context Categorization System

In order to validate the proposed categorization system of medical contexts, we have used test dataset and calculated the accuracy in the form of F-score. Table 6 shows the extracted categories of medical contexts and a statistical comparison with manually labeled test dataset for medical context categories.

The presented result shows the importance of the proposed eleven categories of medical contexts to discover the overall conceptual information from the corpus. We have also observed that the category of the context provides an overall intuition whereas category of each concept of the context indicates specific fundamental class. Finally, we can conclude that the proposed systems may help to develop various medical applications like medical concept network, annotation system, and recommendation system in healthcare.

5 Conclusion and Future Scope

Category refers the broadest fundamental classes of any information extraction system especially in medical domain to identify the subjective information from the corpus. It also helps to represent structured corpus from the largely available unstructured and semi-structured corpora. So, in this paper, we are motivated to design the categorization system for medical concepts and contexts using supervised classifiers in the presence of a domain-specific lexicon namely WME 3.0.

The lexicon assists in preparing the experimental dataset and extracting various features from the dataset to build and validate the proposed categorization systems. The concept categories show the subjective information for each of the concepts whereas pair-based context categories indicate the overall conceptual information for the medical context. We are able to identify five types of concept categories as *diseases* and *symptoms* and eleven types of context categories like *disease-symptom* by the proposed systems individually. In order to build the category assignment system for medical concepts, we have amalgamated two classifiers namely Naïve Bayes and Logistic Regression. We have also conducted an ablation study to prove the accuracy of the system.

Thereafter, we have described how this system helps to design context categorization systems. In future, we will try to introduce more knowledge-based features to improve the accuracy of both of the systems. We will also focus on to design various applications to support expert and non-expert groups in their respective applications.

References

1. Abacha, A. B., Zweigenbaum, P.: Automatic extraction of semantic relations between medical entities: a rule based approach. Journal of Biomedical Semantics, vol. 2 (2011) doi: 10.1186/2041-1480-2-S5-S4
2. Abacha, A. B., Zweigenbaum, P.: A hybrid approach for the extraction of semantic relations from MEDLINE abstracts. In: International Conference on Intelligent Text Processing and Computational Linguistics, pp. 139–150 (2011) doi: 10.1007/978-3-642-19437-5_11
3. Basili, R., Pazienza, M. T., Vindigni, M.: Corpus-driven unsupervised learning of verb subcategorization frames. In: Congress of the Italian Association for Artificial Intelligence, pp. 159–170 (1997) doi: 10.1007/3-540-63576-9_105
4. Bodenreider, O.: The unified medical language system (UMLS): Integrating biomedical terminology. Nucleic Acids Research, vol. 32, pp. D267–D270 (2004) doi: 10.1093/nar/gkh061
5. Borthwick, A., Sterling, J., Agichtein, E., Grishman, R.: Exploiting diverse knowledge sources via maximum entropy in named entity recognition. In: Proceedings of the Sixth Workshop on Very Large Corpora, vol. 182 (1998)
6. Brown, S. H., Elkin, P. L., Rosenbloom, S. T., Fielstein, E., Speroff, T.: eQuality for all: Extending automated quality measurement of free text clinical narratives. In: AMIA Annual Symposium Proceedings, vol. 2008, pp. 71–75 (2008)
7. Cambria, E.: Affective computing and sentiment analysis. IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107 (2016) doi: 10.1109/MIS.2016.31
8. Cambria, E., Hussain, A., Durrani, T., Havasi, C., Eckl, C., Munro, J.: Sentic computing for patient centered applications. In: IEEE 10th International Conference on Signal Processing Proceedings, pp. 1279–1282 (2010) doi: 10.1109/ICOSP.2010.5657072
9. Cambria, E., Hussain, A., Eckl, C.: Bridging the Gap between structured and unstructured healthcare data through semantics and sentics. In: Proceedings of the ACM Web Science Conference, pp. 1–4 (2011)
10. Cambria, E., Poria, S., Hazarika, D., Kwok, K.: SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In: The Thirty-Second AAAI Conference on Artificial Intelligence, vol. 32 (2018) doi: 10.1609/aaai.v32i1.11559
11. Chaturvedi, I., Ragusa, E., Gastaldo, P., Zunino, R., Cambria, E.: Bayesian network based extreme learning machine for subjectivity detection. Journal of The Franklin Institute, vol. 355, no. 4, pp. 1780–1797 (2018) doi: 10.1016/j.jfranklin.2017.06.007
12. Eklund, A. M.: Relational annotation of scientific medical corpora. In: LOUHI 2011 Third International Workshop on Health Document Text Mining and Information Analysis, pp. 27–34 (2011)
13. Embarek, M., Ferret, O.: Learning patterns for building resources about semantic relations in the medical domain. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation (2008)
14. Esuli, A., Sebastiani, F.: SENTIWORDNET: A publicly available lexical resource for opinion mining. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation, vol. 6, pp. 417–422 (2006)

15. Fellbaum, C.: WordNet: An electronic lexical database. The MIT Press (2000) doi: 10.7551/mitpress/7287.001.0001
16. Franzén, K., Eriksson, G., Olsson, F., Asker, L., Lidén, P., Cöster, J.: Protein names and how to find them. International Journal of Medical Informatics, vol. 67, no. 1, pp. 49–61 (2002) doi: 10.1016/s1386-5056(02)00052-7
17. Frunza, O., Inkpen, D.: Extraction of disease-treatment semantic relations from biomedical sentences. In: Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, pp. 91–98 (2010)
18. Kim, Y., Riloff, E., Hurdle, J. F.: A study of concept extraction across different types of clinical notes. In: AMIA Annual Symposium Proceedings, vol. 2015, pp. 737–746 (2015)
19. Lee, C. H., Khoo, C., Na, J. C.: Automatic identification of treatment relations for medical ontology learning: An exploratory study. In: Proceedings of the Eighth International ISKO Conference, pp. 245–250 (2004)
20. Lee, C. H., Na, J. C., Khoo, C.: Ontology learning for medical digital libraries. In: International Conference on Asian Digital Libraries, vol. 2911, pp. 302–305 (2003) doi: 10.1007/978-3-540-24594-0_29
21. Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., Hurdle, J. F.: Extracting information from textual documents in the electronic health record: A review of recent research. Yearbook of Medical Informatics, vol. 17, no. 1, pp. 128–144 (2008) doi: 10.1055/s-0038-1638592
22. Mondal, A., Cambria, E., Das, D., Bandyopadhyay, S.: MediConceptNet: An affinity score based medical concept network. In: Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, pp. 335–340 (2017)
23. Mondal, A., Cambria, E., Feraco, A., Das, D., Bandyopadhyay, S.: Auto-categorization of medical concepts and contexts. In: IEEE Symposium Series on Computational Intelligence, pp. 1–7 (2017) doi: 10.1109/SSCI.2017.8285253
24. Mondal, A., Chaturvedi, I., Das, D., Bajpai, R., Bandyopadhyay, S.: Lexical resource for medical events: A polarity based approach. In: 2015 IEEE International Conference on Data Mining Workshop, pp. 1302–1309 (2015) doi: 10.1109/ICDMW.2015.170
25. Mondal, A., Das, D., Cambria, E., Bandyopadhyay, S.: WME: Sense, polarity and affinity based concept resource for medical events. In: Proceedings of the Eighth Global WordNet Conference, pp. 243–248 (2016)
26. Mondal, A., Das, D., Cambria, E., Bandyopadhyay, S.: WME 3.0: An enhanced and validated lexicon of medical concepts. In: Proceedings of the Ninth Global WordNet Conference, pp. 10–16 (2018)
27. Mondal, A., Satapathy, R., Das, D., Bandyopadhyay, S.: A hybrid approach based sentiment extraction from medical context. In: Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology, pp. 35–40 (2016)
28. Neves, M.: An analysis on the entity annotations in biological corpora. F1000Research, (2014) doi: 10.12688/f1000research.3216.1
29. Pakhomov, S. V., Coden, A., Chute, C. G.: Developing a corpus of clinical notes manually annotated for part-of-speech. International journal of medical informatics, vol. 75, no. 6, pp. 418–429 (2006) doi: 10.1016/j.ijmedinf.2005.08.006
30. Rink, B., Harabagiu, S., Roberts, K.: Automatic extraction of relations between medical concepts in clinical texts. Journal of the American Medical Informatics Association, vol. 18, no. 5, pp. 594–600 (2011) doi: 10.1136/amiajnl-2011-000153
31. Rosario, B., Hearst, M. A.: Multi-way relation classification: Application to protein-protein interactions. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 732–739 (2005)

32. Shukla, R. S., Yadav, K. S., Abbas, S. T., Haseen, F.: An efficient mining of biomedical data from hypertext documents via NLP. In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, vol. 327, pp. 651–658 (2015) doi: 10.1007/978-3-319-11933-5_73
33. Smith, B., Fellbaum, C.: Medical WordNet: a new methodology for the construction and validation of information resources for consumer health. In: Proceedings of the 20th international conference on Computational Linguistics, pp. 371–382 (2004) doi: 10.3115/1220355.1220409
34. South, B. R., Mowery, D., Suo, Y., Leng, J., Ferrández, O., Meystre, S. M., Chapman, W. W.: Evaluating the effects of machine pre-annotation and an interactive annotation interface on manual de-identification of clinical text. Journal of Biomedical Informatics, vol. 50, pp. 162–172 (2014) doi: 10.1016/j.jbi.2014.05.002
35. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational linguistics, vol. 37, no. 2, pp. 267–307 (2011) doi: 10.1162/COLI_a_00049
36. Uzuner, Ö., South, B. R., Shen, S., DuVall, S. L.: 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association, vol. 18, no. 5, pp. 552–556 (2011) doi: 10.1136/amiajnl-2011-000203
37. Xia, F., Yetisgen-Yildiz, M.: Clinical corpus annotation: Challenges and strategies. In: Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM'2012) in conjunction with the International Conference on Language Resources and Evaluation (2012)
38. Yao, L., Sun, C. J., Wang, X. L., Wang, X.: Relationship extraction from biomedical literature using maximum entropy based on rich features. In: 2010 International Conference on Machine Learning and Cybernetics, vol. 6, pp. 3358–3361 (2010) doi: 10.1109/ICMLC.2010.5580680