WME 3.0:
An Enhanced and Validated Lexicon of Medical Concepts

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

[1]Department of Computer Science and Engineering   [2]School of Computer Science and Engineering
Jadavpur University, Kolkata, India   Nanyang Technological University, Singapore
[1]`anupam@sentic.net,`[1]`dipankar.dipnil2005@gmail.com`
[2]`cambria@ntu.edu.sg,`[1]`sivaji_cse_ju@yahoo.com`

Abstract

Information extraction in the medical domain is laborious and time-consuming due to the insufficient number of domain-specific lexicons and lack of involvement of domain experts such as doctors and medical practitioners. Thus, in the present work, we are motivated to design a new lexicon, WME 3.0 (WordNet of Medical Events), which contains over 10,000 medical concepts along with their part of speech, gloss (descriptive explanations), polarity score, sentiment, similar sentiment words, category, affinity score and gravity score features. In addition, the manual annotators help to validate the overall as well as individual category level of medical concepts of WME 3.0 using Cohen's Kappa agreement metric. The agreement score indicates almost correct identification of medical concepts and their assigned features in WME 3.0.

1 Introduction

In the clinical domain, the representation of a lexical resource is treated as a crucial and contributory task because of handling several challenges. The challenges are the identification of medical concepts, their categories and relations, disambiguation of polarities, recognition of semantics whereas the scarcity of structured clinical texts doubles the challenges. In the last few years, several researchers were involved in developing various domain-specific lexicon such as Medical WordNet and UMLS (Unified Medical Language System) to cope up with such challenges. These lexicons help to bridge the gap between medical experts such as doctors or medical practitioners and non-experts such as patients (Cambria et al., 2010a; Cambria et al., 2010b).

However, medical text is in general unstructured since doctors do not like to fill forms and prefer free-form notes of their observations. Hence, a lexical design is difficult due to lack of any prior knowledge of medical terms and contexts. Therefore, we are motivated to enhance a medical lexicon namely WordNet of Medical Events (WME 2.0) which helps to identify medical concepts and their features. In order to enrich this lexicon, we have employed various well-known resources like conventional WordNet, SentiWordNet (Esuli and Sebastiani, 2006), SenticNet (Cambria et al., 2016), Bing Liu (Liu, 2012), and Taboada's Adjective list (Taboada et al., 2011) and a preprocessed English medical dictionary[1] on top of WME 1.0 and WME 2.0 lexicons (Mondal et al., 2015; Mondal et al., 2016). WME 1.0 contains 6415 number of medical concepts and their glosses, POS, polarity scores, and sentiment. Thereafter, Mondal et. al., (2016) enhanced WME 1.0 by adding few more features as affinity score, gravity score, and SSW to the medical concepts and presented as WME 2.0. The affinity and gravity scores present the hidden link between the pair of medical concepts and the concept with the various source of glosses respectively. SSW of a medical concept refers the similar sentiment words (SSW) which follow the common sentiment property.

In the current research, we have focused on enriching WME 2.0 with more number of medical concepts and including an additional feature i.e medical category. In order to develop such updated version of WME namely WME 3.0, we have taken the help of WME 1.0 and WME 2.0. We have also noticed that the previous versions of WMEs are unable to extract knowledge-based information such as the category of the medical concepts and its coverage is also lower.

---

[1]http://alexabe.pbworks.com/f/Dictionary+of+Medical+Terms+4th+Ed.-+(Malestrom).pdf

Therefore, we have enhanced the number of medical concepts as well as add category feature on top of WME 2.0. The current version, WME 3.0 contains 10,186 number of medical concepts and their category, POS, gloss, sentiment, polarity score, SSW, affinity and gravity scores. For example, WME 3.0 lexicon presents the properties of a medical concept say **amnesia** as of category (*disease*), POS (*noun*), 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.*), sentiment (*negative*), polarity score (*-0.375*), SSW (*memory_loss, blackout, fugue, stupor*), affinity score (*0.429*) and gravity score (*0.170*).

Moreover, to enhance and validate lexicon with the newly added medical concepts and categories, we have summarized our contributions as follows.

(a) Enriching the number of medical concepts in the existing lexicon, WME 2.0: In order to meet up this issue, we have employed a preprocessed English medical dictionary[2] and various well-defined lexicons such as SentiWordNet, SenticNet, and MedicineNet etc. They helped to enhance the number of medical concepts of the proposed lexicon.

(b) Overall validation of the current lexicon: To resolve the issue, we have taken the help of two manual annotators as medical practitioners. The annotators provided agreement scores that are processed using Cohen's Kappa and obtained a κ score which assists in validating the overall lexicon as well as the individual features of WME 3.0 (Viera et al., 2005).

(c) Evaluate various individual feature of the medical concepts: In order to extract the subjective and knowledge-based features, we have applied our evaluation scripts on the mentioned resources. The scripts assist in identifying the affinity and gravity scores as feature values for the concepts. Also, the resources are used to assign the SSW as semantics and glosses for the concepts. On the other hand, a supervised classifier helps to add the category feature in the proposed lexicon.

The remainder of the paper is organized as follows: Section 2 presents the related works for building a medical lexicon; Section 3 and Section 4 describe the previous versions of WMEs like WME 1.0 and WME 2.0 and the development

steps of WME 3.0; Section 5 discusses the validation process of the proposed lexicon; finally, Section 6 illustrates the concluding remarks and future scopes of the research.

2 Background

Biomedical information extraction is treated as one of the challenging research tasks as it deals with available medical corpora that are either unstructured or semi-structured. Hence, a domain-specific lexicon becomes an essential component to convert a structured corpus from the unstructured corpus (Borthwick et al., 1998). Also, it helps in extracting the subjective and conceptual information related to medical concepts from the corpus. Various researchers have tried to build various ontologies and lexicons such as UMLS, SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms), MWN (Medical WordNet), SentiHealth, and WordNet of Medical Events (WME 1.0 and WME 2.0) etc. in the domain of healthcare (Miller and Fellbaum, 1998; Smith and Fellbaum, 2004; Asghar et al., 2016; Asghar et al., 2014). UMLS helps to enhance the access to biomedical literature by facilitating the development of computer systems that understand biomedical language (Bodenreider, 2004). SNOMED-CT is a standardized, multilingual vocabulary that contains clinical terminologies and assists in exchanging the electronic healthcare information among physicians (Donnelly, 2006).

Furthermore, Fellbaum and Smith (2004) proposed Medical WordNet (MWN) with two sub-networks e.g., Medical FactNet (MFN) and Medical BeliefNet (MBN) for justifying the consumer health. The MWN follows the formal architecture of the Princeton WordNet (Fellbaum, 1998). On the other hand, MFN aids in extracting and understanding the generic medical information for non-expert groups whereas MBN identifies the fraction of the beliefs about the medical phenomena (Smith and Fellbaum, 2004). Their primary motivation was to develop a network for medical information retrieval system with visualization effect. SentiHealth lexicon was developed to identify the sentiment for the medical concepts (Asghar et al., 2016; Asghar et al., 2014). WME 1.0 and WME 2.0 lexicons were designed to extract the medical concepts and their related linguistic and sentiment features from the corpus (Mondal et al., 2015; Mondal et al., 2016).

---

[2] http://alexabe.pbworks.com/f/Dictionary+of+Medical+Terms+4th+Ed.-+(Malestrom).pdf

These mentioned ontologies and lexicons assist in identifying the medical concepts and their sentiments from the corpus but unable to provide the complete knowledge-based information of the concepts. Hence, in the current work, we are motivated to design a full-fledged lexicon in healthcare which provides the linguistic, sentiment, and knowledge-based features together for the medical concepts.

3 Attempts for WordNet of Medical Events

In healthcare, a domain-specific lexicon is required for identifying the conceptual and knowledge-based information such as category, gloss, semantics, and sentiment of the medical concepts from the clinical corpora (Cambria, 2016). We have borrowed the knowledge from a domain-specific lexicon namely WordNet of Medical Events (WME) with its two different versions such as WME 1.0 and WME 2.0. These versions are distinguished according to the versatility and variety of medical concepts and their features.

3.1 WME 1.0

WME 1.0 contains 6415 numbers of medical concepts and their linguistic features such as gloss, parts of speech (POS), sentiment and polarity score (Mondal et al., 2015). The gloss and POS represent the descriptive definition and linguistic nature of the medical concepts whereas the sentiment and polarity score refer the classes as positive, negative, and neutral and their corresponding strength (+1) and weakness (-1). The resource was prepared by employing the trial and training datasets of SemEval-2015 Task-6[3] which initially contains only 2479 medical concepts. Thereafter, the extracted concepts were updated using WordNet and a preprocessed English medical dictionary as mentioned earlier for enriching the number of concepts and identifying gloss and POS of them. However, sentiment and polarity scores were added afterwards using sentiment lexicons such as SentiWordNet[4], SenticNet[5], Bing Liu's subjective list[6], and Taboada's adjective list[7] (Cambria et al., 2016; Taboada et al., 2011; Esuli and Sebastiani, 2006).

[3]http://alt.qcri.org/semeval2015/task6/
[4]http://sentiwordnet.isti.cnr.it/
[5]http://sentic.net/downloads/
[6]https://www.cs.uic.edu/
[7]http://neuro.imm.dtu.dk/wiki/

For example, the medical concept **abnormality** appears with the following gloss, POS as noun, negative sentiment and polarity score of -0.25 in WME 1.0.

3.2 WME 2.0

The next version of WME, i.e., WME 2.0, extracts more semantic features of medical concepts (Mondal et al., 2016) and added with the existing features of WME 1.0. While updated WME 2.0 with affinity score, gravity score, and SSW, the number of concepts in WME 2.0 remains same, but the features of each concept are included (Mondal et al., 2016).

Affinity score indicates the strength of a medical concept and its corresponding SSWs by assigning a probability score. SSW of a medical concept presents the SSW shared through their common sentiment property. The affinity score '0' indicates no relation whereas '1' suggests a strong relationship between a pair of concepts. On the other hand, gravity score helps to extract the sentiment relevance between a concept and its glosses. It ranges from -1 to 1 including 0 while '-1' suggests no relation, '0' describes neutral situations of either concept or gloss without sentiment, and '1' indicates strong relations either positive or negative. It is used to prove the knowledge-based relevance between a concept and its gloss. In order to extract the features, the authors used WordNet, SentiWordNet, SenticNet, and a preprocessed English medical dictionary. Figure 1 shows the presentation of WME 2.0 lexicon for a medical concept **abnormality**.

In the present research, we have enriched the number of medical concepts and category feature with WME 2.0 lexicon and presented the enhanced version WME 3.0. The following section discusses the steps of WME 3.0 building.

4 Development of WME 3.0

A large number of daily produced medical corpora and their adaptable natures introduce the difficulty to build a full-fledged medical lexicon in healthcare domain. In order to resolve the issue, we have proposed a new version of WordNet of Medical Events namely WME 3.0. It is observed that WME 3.0 helps to extract more medical concepts and features from the unstructured corpus with respect to the previous version of WME, i.e., WME 2.0.

```
<Concept>
<Title>abnormality</Title>
<Properties>
      <Affinity_score>0.692</Affinity_score>
      <Gloss>An abnormal physical condition resulting from
      defective genes or developmental deficiencies.</Gloss>
      <Gravity_score>0.125</Gravity_score>
      <Polarity_score>-0.25</Polarity_score>
      <POS>Noun</POS>
      <Sentiment>Negative</Sentiment>
      <SSW>Anomaly,Peculiarity,Extraordinariness</SSW>
</Properties>
</Concept>
```

Figure 1: An example of assigned features of a medical concept *abnormality* under WME 2.0 lexicon.

Another 3771 number of medical concepts and an additional category feature were newly added into WME 3.0. Finally, WME 3.0 contains 10,186 medical concepts and their POS, categories, affinity scores, gravity scores, polarity scores, sentiments and SSW. To identify the additional medical concepts, we have employed the conventional WordNet[8] and MedicineNet[9] resource. Thereafter, we have written a script to extract new medical concepts, which are semantically (like common POS as well as sentiment) related with medical concepts of WME 2.0. Besides, SentiWordNet, SenticNet, Bing Liu subjective list, Taboada's adjective list, and previously mentioned preprocessed medical dictionary help to assign all features except category to 3771 medical concepts which were added.

Thereafter, we newly considered four different types of categories namely diseases, drugs, symptoms, and human_anatomy for this research after examining the nature of medical concepts. In WME 3.0, all concepts are tagged with either the above-mentioned four categories or MMT category. MMT represents the miscellaneous medical terms which refer to the uncategorized and unrecognized medical concepts. In order to assign the category to the medical concepts, we have applied a well-known machine learning classifier, Naïve Bayes on top of WME 3.0 driven features. The classifier learns through the manually annotated 2000 medical concepts and their categories. Thereafter, rest of 8186 medical concepts of WME 3.0 were processed by the classifier by predicting the category (Mondal et al., 2017a).

For example, the medical concept *ranitidine* represents the category, *drug* in WME 3.0 lexicon. Table 1 illustrates a comparative analysis and progress reports on WME 1.0, WME 2.0, and WME 3.0 with respect to the coverage of medical concepts, n-gram counts, and other different features such as POS, sentiment, polarity score, affinity score, gravity score, and category.

We have also noticed that the proposed WME 3.0 primarily contains POS as a noun, sentiment as negative, category as disease and drug, and n-gram feature as uni-grams and bi-grams. The observations could help to understand the characteristic of the lexicon and assist in designing various applications viz. medical annotation and concept network systems etc. The lexicon is very much demanding to identify four different types of categories and each medical concepts related gloss from a medical corpus, which presents the difference between WME 3.0 and already established very large scale semantic networks, such as UMLS. Also, the lexicon-driven medical concepts and their features also assist in emulating human thought as a recommendation of medical advice, serving a potential foundation of a higher-order cognitive model under natural language processing (Cambria and Hussain, 2015; Cambria et al., 2011). Finally, the evaluation process of WME 3.0 as overall and its individual feature levels are discussed in the following section.

5   Evaluation

In order to validate our proposed WME 3.0 lexicon, we have conducted the following result analysis. The result shows the agreement between two manual annotators to explain the acceptance

---

Features		WME 1.0	WME 2.0	WME 3.0
No. of Concepts		6415	6415	10186
n-grams	*Uni-gram*	2956	2956	3722
	Bi-gram	2837	2837	3866
	Tri-gram	622	622	1762
POS	*Noun*	4248	4248	7677
	Verb	2056	2056	2352
	Adjective	111	111	157
Sentiment and Polarity_score	*Positive (>= 1)*	2800	2800	3227
	Negative (< 1)	3615	3615	6959
Affinity_score	*0 to 0.5*	-	4325	7177
	0.5 to 1	-	2090	3009
Gravity_score	*less than zero*	-	2320	3783
	equal to zero	-	732	1961
	grater than zero	-	3363	4442
Category	*Disease*	-	-	3243
	Drug	-	-	3390
	Symptom	-	-	1409
	Human_Anatomy	-	-	227
	MMT	-	-	1917

Table 1: [Color online] A comparative statistics for various features of medical concepts present in WME 1.0 (Blue), WME 2.0 (Green), and WME 3.0 (Yellow).

of overall lexicon as well as its individual features. The agreement has been calculated using Cohen's Kappa coefficient score κ which is defined in Equation 1 (Viera et al., 2005).

$$\kappa = \frac{Pr_a - Pr_e}{1 - Pr_e}, \qquad (1)$$

where Pr_a is the observed proportion of full agreement between two annotators. In addition, Pr_e is the proportion expected by a chance which indicates a kind of random agreement between the annotators.

5.1 Overall Validation of WME 3.0

WME 3.0 has been validated by two manual annotators, where the annotators are medical practitioners. The annotators have verified both medical concepts and their category, POS, gloss, affinity score, gravity score, polarity score, SSW, and sentiment features and presented as a number of yes (agreed) and number of no (disagreed) values. Table 2 indicates the values provided by both of the annotators in terms of agreement-based scores. The scores produced 0.79 κ score using equation 1. The κ score shows significantly approved result for WME 3.0 lexicon.

No. of Concepts: 10186		Annotator-1	
		Yes	No
Annotator-2	**Yes**	8629	189
	No	285	1083

Table 2: An agreement analysis between two annotators to validate medical concepts and their all features under WME 3.0.

5.2 Individual Feature based Validation of WME 3.0

On the other hand, the same annotators also assist in validating the individual feature of WME 3.0 with respect to the medical concepts. Hence, we have split the proposed lexicon into five parts where each of the parts contains the medical concepts and its corresponding primary features viz. category, POS, gloss, SSW, and sentiment individually. We have not considered rest of the three features namely affinity, gravity, and polarity scores of WME 3.0 because these features were derived from the above-mentioned five primary features. Thereafter, the annotators help to validate the five parts by counting the number of yes (agreed) and no (disagreed) individually. The provided agreement counts are processed with Equation 1 and get 0.89, 0.91, 0.88, 0.82, and 0.92 κ scores for category, POS, gloss, SSW, and sentiment, respectively.

The κ scores prove the usefulness and quality of individual features of the medical concepts for WME 3.0. Table 3 shows the agreement statistics between two annotators for validating the features of WME 3.0 lexicon.

No. of Concepts: 10186		Annotator-1		κ score
		Yes	No	
Annotator-2	Category	Yes	8778	93	
		No	161	1154	0.89
	POS	Yes	9229	52	
		No	92	813	0.91
	Gloss	Yes	8805	97	
		No	172	1112	0.88
	SSW	Yes	8767	137	
		No	256	1026	0.82
	Sentiment	Yes	8727	67	
		No	124	1268	0.92

Table 3: An agreement analysis between two annotators to validate category, POS, Gloss, SSW, and Sentiment features of medical concepts of WME 3.0.

We have analyzed the agreement scores for the features of WME 3.0. It is found that all the features of medical concepts are quite correctly labeled in the lexicon as presented in Table 3. We have also observed that the disagreement has been occurred due to the conceptual mismatch between two annotators or place of the usage of a few medical concepts for each of the features.

For example, the medical concept ***blood_clot*** is tagged with either *symptom* or *disease* category. In case of POS, the medical concept ***abnormality*** is either labeled as an *adjective* or a *noun* whereas ***menstrual_cycle*** refers *positive* or *negative* sentiment. Such types of disagreements are treated as very difficult task for the contextual behavior of medical corpora.

Besides, we have studied each type of the categories such as disease, symptom, and drug etc. to justify their presence in WME 3.0 lexicon. The annotators again help to validate each of the assigned categories using agreement analysis as shown in Table 4. The supplied agreement counts have been applied on Equation 1 and we found 0.89, 0.87, 0.88, 0.90, and 0.91 κ scores for disease, symptom, drug, human_anatomy, and MMT categories, respectively.

Finally, we can conclude that, WME 3.0 lexicon assists in increasing the coverage of the medical concepts as well as features and may be pre-

No. of Concepts		Annotator-1		κ score
		Yes	No	
Annotator-2	Disease (3243)	Yes	2794	31	
		No	51	367	0.89
	Symptom (1409)	Yes	1214	14	
		No	26	155	0.87
	Drug (3390)	Yes	2922	34	
		No	53	381	0.88
	Human_anatomy (227)	Yes	196	2	
		No	3	26	0.90
	MMT (1917)	Yes	1652	12	
		No	28	225	0.91

Table 4: An agreement analysis between two annotators to validate individual categories of WME 3.0.

sented as a full-fledged lexicon in the healthcare domain. Also, the lexicon can take a crucial role to design various applications such as medical annotation, concept network, and relationship identification system in healthcare (Mondal et al., 2017b).

6 Conclusion and Future Work

The present task has been motivated to enrich a medical lexicon with additional medical concepts and a feature called category in WME 3.0. In order to prepare the current version, we have employed previous two versions of WME viz. WME 1.0 and WME 2.0 along with various well-defined lexicons and a machine learning classifier. WME 3.0 contains 10,186 medical concepts and eight different types of useful features such as category and gloss etc.

In addition, we have also validated WME 3.0 from two different aspects, namely overall evaluation and usefulness of individual feature with the help of two manual annotators. The annotators provided agreement scores that were processed using Cohen's kappa agreement analysis. Finally, the κ scores showed the importance of WME 3.0 in healthcare. In future, we will attempt to enhance WME 3.0 with more number of medical concepts as well as syntactic and semantic features for improving the coverage and quality.

References

Muhammad Z Asghar, Aurangzeb Khan, Fazal M Kundi, Maria Qasim, Furqan Khan, Rahman Ullah, and Irfan U Nawaz. 2014. Medical opinion lexicon: an incremental model for mining health reviews. *International Journal of Academic Research*, 6(1):295–302.

Muhammad Zubair Asghar, Shakeel Ahmad, Maria Qasim, Syeda Rabail Zahra, and Fazal Masud Kundi. 2016. SentiHealth: creating health-related sentiment lexicon using hybrid approach. *SpringerPlus*, 5(1):1139.

Olivier Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270.

Andrew Borthwick, John Sterling, Eugene Agichtein, and Ralph Grishman. 1998. Exploiting diverse knowledge sources via maximum entropy in named entity recognition. In *Proc. of the Sixth Workshop on Very Large Corpora*, volume 182.

Erik Cambria and Amir Hussain. 2015. *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Springer, Cham, Switzerland.

Erik Cambria, Amir Hussain, Tariq Durrani, Catherine Havasi, Chris Eckl, and James Munro. 2010a. Sentic computing for patient centered applications. In *IEEE 10th International Conference on Signal Processing Proceedings*, pages 1279–1282. IEEE.

Erik Cambria, Amir Hussain, Catherine Havasi, Chris Eckl, and James Munro. 2010b. Towards crowd validation of the UK national health service. In *WebSci*, Raleigh.

Erik Cambria, Thomas Mazzocco, Amir Hussain, and Chris Eckl. 2011. Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space. In D Liu, H Zhang, M Polycarpou, C Alippi, and H He, editors, *Advances in Neural Networks*, volume 6677 of *Lecture Notes in Computer Science*, pages 601–610, Berlin. Springer-Verlag.

Erik Cambria, Soujanya Poria, Rajiv Bajpai, and Björn W Schuller. 2016. SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives. In *COLING*, pages 2666–2677.

Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107.

Kevin Donnelly. 2006. SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in health technology and informatics*, 121:279.

Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422. Citeseer.

Christiane Fellbaum. 1998. *WordNet*. Wiley Online Library.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

George Miller and Christiane Fellbaum. 1998. WordNet: An electronic lexical database.

Anupam Mondal, Iti Chaturvedi, Dipankar Das, Rajiv Bajpai, and Sivaji Bandyopadhyay. 2015. Lexical Resource for Medical Events: A Polarity Based Approach. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1302–1309. IEEE.

Anupam Mondal, Dipankar Das, Erik Cambria, and Sivaji Bandyopadhyay. 2016. WME: Sense, Polarity and Affinity based Concept Resource for Medical Events. *Proceedings of the Eighth Global WordNet Conference*, pages 242–246.

Anupam Mondal, Erik Cambria, Dipankar Das, and Sivaji Bandyopadhyay. 2017a. Auto-categorization of medical concepts and contexts. In *IEEE Symposium Series on Computational Intelligence (SSCI 2017), Honolulu, Hawaii, USA*.

Anupam Mondal, Erik Cambria, Dipankar Das, and Sivaji Bandyopadhyay. 2017b. MediConceptNet: An Affinity Score Based Medical Concept Network. In *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2017, Marco Island, Florida, USA, May 22-24, 2017.*, pages 335–340.

Barry Smith and Christiane Fellbaum. 2004. 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*, page 371. Association for Computational Linguistics.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

Anthony J Viera, Joanne M Garrett, et al. 2005. Understanding inter-observer agreement: the kappa statistic. *Fam Med*, 37(5):360–363.