Combining Sentiment Lexicons of Arabic Terms

Full paper

Nabeela Altrabsheh
Qatar University
nabeela@qu.edu.qa

Mazen El-Masri
Qatar University
mazen.elmasri@qu.edu.qa

Hanady Mansour
Qatar University
hanadyma@qu.edu.qa

Abstract

Lexicons are dictionaries of sentiment words and their matching polarity. Some comprise words that are numerically scored based on the degree of positivity/negativity of the underlying sentiments. The ranges of scores differ since each lexicon has its own scoring process. Others use labelled words instead of scores with polarity tags (i.e., positive/negative/neutral). Lexicons are important in text mining and sentiment analysis which compels researchers to develop and publish them. Larger lexicons better train sentiment models thereby classifying sentiments in text more accurately. Hence, it is useful to combine the various available lexicons. Nevertheless, there exist many duplicates, overlaps and contradictions between these lexicons. In this paper, we define a method to combine different lexicons. We used the method to normalize and unify lexicon items and merge duplicated lexicon items from twelve lexicons for (in)formal Arabic. This resulted in a coherent Arabic sentiment lexicon with the largest number of terms.

Keywords

Lexicons, Sentiment Analysis, Arabic

Introduction

Sentiment analysis in simple form is to classify opinions into polarity (Liu, B., 2012). The sentiment polarity of a word can either be positive (e.g., happy), negative (e.g., upset), or neutral (e.g., tall) (Liu, B., 2012). There are three main approaches to sentiment analysis. They are lexicon-based, machine learning, and hybrid approach.

The first approach is lexicon-based method where a lexicon is used to score the sentence or label it into polarity (Taboada et al., 2011). A lexicon is a dictionary which contain words and their matching polarity label or score (Taboada et al., 2011). These lexicons are created from large amounts of labelled data (i.e., opinions) (Taboada et al., 2011). The word occurrences according to each label is counted to create the lexicon. Lexicon terms are scored differently in the extant literature. Some use mathematical equations to give a score words into strong positive (e.g., thrilled) or weak positive (e.g., pleased) (Al-Twairesh et al., 2016; Mohammad et al., 2016). Others use labelled words instead of scores with simple polarity tags (i.e., positive, negative, and neutral) (El Sahar et al., 2014; El-Beltagy and Ali, 2013). Lexicons can be general or domain-specific. The domain-specific lexicons are created from data that is from a particular domain such as education or movies. The general lexicons are data collected from several domains or opinions that do not fall into one domain. Whether general or domain-specific, lexicons must regularly get updated. Opinions are changeable and new words continually appear (Maurer and High,1980). Moreover, the usefulness of lexicons is limited to the words that are in the lexicon. Therefore, it is important to create larger lexicons with more sentiment words.

The second approach is the machine learning method which uses different classifiers to classify sentences into polarity. The machine learning method is more common than the lexicon-based. It usually requires labelled data. The new data will be classified according to algorithms and the training data input. There are many classifiers that can be used to detect the polarity. Some of the classifiers perform better with one domain than others. For example, the Support Vector Machine (SVM) Classifier performed best with the education domain for students' feedback in Alabaster et al. (2014).

The last approach is the combined or hybrid. This approach use two or more approaches –usually the lexicon and machine learning approaches combined – to analyse sentiment. According to many studies, the combined approach usually gives more accurate results (Prabowo and Thelwall, 2009).

Sentiment analysis models depend on the language of the opinions. Most of the sentiment analysis research has been on English language and only few studies explored sentiment analysis of Arabic language. Some of those studies have made their lexicons publically available. Yet, these lexicons are a mix of Modern Standard Arabic (MSA) and Dialectical Arabic (DA). Each of these lexicons are scored differently. In Arabic language, there exist different dialects according to the region and country. The main dialects according to each region are: Gulf, Levantine, Egypt, Yemeni, Moroccan, and Iraqi (Darwish,2012). This makes sentiment analysis more complex as there exist a large variety of vocabulary for each dialect. Therefore, there is a need of lexicons consisting of different dialects including standard Arabic.

The available lexicons contain data overlaps and contradictions due to the time or domain the data is collected in. However, it is important to find ways to use the available resources and lexicons together. Combining lexicons available online allows us to get more accurate analysis of the sentiment. Cho et al. (2013) found that combining different lexicons can lead to a better accuracy. Combining Arabic lexicons allows the sentiment analysis models to cover both MSA and DA words. While there exist some studies who have combined lexicons by removing duplicates, our study explores a new method. Our contribution is to combine between different lexicons and to create a large lexicon that can be used for sentiment analysis purposes. To the best of our knowledge, this is the first attempt to combine lexicons in Arabic language. We have used a mathematical formula and created a new algorithm to combine sentiment lexicons.

Related Work

There are only a few examples in the literature that have explored merging lexicons in general. Some examples are Das and Bandyopadhyay (2010a), and Das and Bandyopadhyay (2010b), Ohana (2011), Cho et al. (2013), Filho et al.(2013), and Emerson and Declerck (2014). These will be discussed in the following paragraphs.

Das and Bandyopadhyay (2010a) merged three lexicons: Charles Philip Brown English-Telugu Dictionary, Aksharamala English-Telugu Dictionary and English-Telugu Dictionary. They removed duplicates automatically by choosing the first existence. This resulted in 112310 unique entries. However, removing only one entry could lead to errors and inaccurate results, as they did not take into consideration the knowledge brought in from the other lexicons.

In another study, Das and Bandyopadhyay (2010b) merged between SentiWordNet and Subjectivity Word List lexicons. They removed duplicates by choosing the first existence and found that 64% of words are common in the Subjectivity Word List and SentiWordNet. The final merged sentiment lexicon contained 14,135 words.

Ohana (2011) explored different lexicons on several domains. They integrate four sentiment lexicons: General Inquirer, the Subjectivity Clues, SentiWordNet, and Moby. They used General Inquirer as a baseline lexicon for their comparisons. The Subjectivity Clues lexicon was close on size and agreement with the General Inquirer. On the other hand, SentiWordNet and Moby were high on disagreement. The researchers only take the agreement factor and discard the disagreements. Comparing all lexicons with the General Inquirer lexicon could be a weak approach because some of the other lexicons may be more accurate for a certain topic. Additionally, terms could be domain specific which may be captured more in other lexicons.

Cho et al. (2013) created a new method of merging multiple lexicons. They used seven sentiment lexicons: AFINN, SentiSense, Micro-WNOp, WordNet-Affect, Opinion Lexicon, Subjectivity Lexicon and General Inquirer. They labelled the lexicons in a scale of positive (1) and negative (-1) by human judges. If a term occurred in a ratio of 9 times negative and 1 time positive in one lexicon and 9 times negative and 3 times positive in another they would average the difference into 2 for the positive. Some of the words were switched into another polarity according to the domain. Some examples of these were “conspiracy”, “horror”, and “tragic” which were originally negative but switched into positive polarity. Their final merged lexicon consisted of 12,114 word entries. One limitation to this research is that they did not take into consideration the existence of equals (i.e., terms with an equal positive and negative rate). This approach is a weak approach as the word could have equally positive and negative.

Filho et al.(2013) combined 3 Brazilian lexicons with certain rules which was firstly discard the neutral terms. Secondly when a word had more than one polarity, they chose the first polarity. For example, if a word was both positive and negative. They would pick positive. They then found that their lexicons had around 74 to 97% agreement. This approach is a weak approach as the word could have both a positive and negative polarity.

Emerson and Declerck (2014) used a Bayesian probabilistic model to merge several sentiment lexicons. They use Clematide and Klenner, SentimentWortschatz, GermanSentiSpin, GermanPolarityClues, and MLSA. They normalize scores by multiplying them by a constant factor for each value. Then they sum the scores. They created a new merged lexicon which is publicly available called SentiMerge. This research was the only research that explored merging scored lexicons using normalization and a probabilistic model. Their lexicon included scores and weights to the polarity score. This research only added weights to the lexicon scores rather than change the scores themselves. This approach could be improved by changing the scores without the complication of weights.

To summarise our findings of previous research, we found that it is essential to take into consideration all lexicon inputs when duplicates occur. Also, none of these lexicons mentioned what they do when duplicates are of equal polarity (i.e., term’s score is equal in two polarities). It is important to find ways to address equal polarity occurrences. In this paper, we address weaknesses in other research and find a novel way to combine lexicons. To the best of our knowledge there is no research exploring combining Arabic sentiment lexicons. Exploring methods to combine Arabic lexicons is important because there a wide variety of dialects, roots and duplications.

Data Collection

Our data was collected from twelve lexicons (See Table 1). Seven of these lexicons were dialectical and five were MSA. Six of these lexicons were scored and six were labelled. The smallest lexicon contained 378 words and the largest contained 225329 words. Only two of these lexicons contained neutral terms. The amount of neutral terms in total were only 571 words. The MPQA contained a label named both which is when the positive equals the negative. In the paragraphs below we briefly describe each of these lexicons. In Table 1, Max and Min represent the maximum and minimum of the polarity score in each of the lexicons.

Lexicon	Source	Polarity Type	Size	Max	Min	Positive	Negative	Neutral	Dialectical/ MSA
AntiSenti	Al-Twairesh et al. (2016)	Scores	225329	7.071	-7.78	116475	108854	-	Dialectical
Arabic Emotion Lexicon	Mohammad et al. (2016)	Scores	43229	7.05	-5.606	22959	20339	1	Dialectical
Arabic Hashtag Lexicon	Mohammad et al. (2016)	Scores	22004	20	-8.409	13158	8846	-	Dialectical
Arabic_Hashtag_Dialetical	Mohammad et al. (2016)	Scores	20125	11	-5.877	11946	8179	-	Dialectical
Sentiment Analysis colloquial	El-Makky et al. (2014)	Labels	8871	-	-	2455	6412	-	Dialectical
Slang Lexicon	El Sahar et al. (2014)	Labels	378	-	-	176	202	-	Dialectical
unWeightedOMLexicon	El-Beltagy and Ali (2013)	Labels	4392	-	-	3537	855	-	Dialectical
General Inquirer	-----	Labels	4206	-	-	1915	2291	-	MSA
Bing Liu	Mohammed (2016).	Labels	6789	-	-	2006	4783	-	MSA
MPQA	Mohammed (2016).	Labels	8189	-	-	2718	4901	570	MSA
NRC Emoticon	Mohammed (2016).	Scores	26740	5	-4.999	15210	11530	-	MSA
NRC Hashtags	Mohammed (2016).	Scores	32582	7.526	-6.925	18341	14241	-	MSA

Table 1: Existing Lexicons of the Arabic Language

AntiSenti lexicon was collected from twitter by Al-Twairesh et al. (2016). They used emoticons and Hashtags to collect the Tweets. They then used MADAMIRA tool and pointwise mutual information to create the lexicon. This was the largest lexicon in our database and was based on dialectical Arabic. This lexicon contained scores.

Arabic Emotion Lexicon, Arabic Hashtag Lexicon, and Arabic Hashtag Dialectical lexicons were created by Mohammad et al. (2016) from Twitter. They were created by measuring the extent to which the words in a tweets corpus co-occurred with a set of seed positive and seed negative terms. This is based on the idea that positive terms co-occur more with positive words and less with negative words; and negative words co-occur more with negative words and less with positive words. This lexicon was a score-based lexicon.

Sentiment Analysis colloquial lexicon was created by El-Makky et al. (2014). It was a merged lexicon from one created previously by El Betagly and Ali (2013). These were based on tweets which were dialectical Arabic. This lexicon was a labelled based lexicon.

Slang Lexicon lexicons was built automatically from Matching tweets to lexico-syntactic patterns. It was created by El Sahar et al. (2014). The UnWeightedOMLexicon lexicon was created by El-Beltagy and Ali (2013). It consists of 4392 entries mostly of Egyptian dialect. Both lexicons were in dialectical Arabic and were labelled-based.

The General Inquirer consist of words available in the Harvard and Lasswell dictionaries. This is a well-known dictionary publically available. We have manually translated it into Arabic text. The amount of words with positive and negative labels were 4206 from 11,788 words. This lexicon is probably one of the most common and used lexicon used in sentiment analysis.

Bing Liu, MPQA, NRC Emoticon, and NRC Hashtags lexicons were translated from different sources by Mohammed (2016). The first lexicon was based on Liu (2010). The second was translated from the MPQA subjectivity lexicon by Wilson et al. (2005). The third one was based on Mohammad and Kiritchenko (2015) and the last one was by Kiritchenko and Mohammad (2014). Bing Liu and MPQA were score-based lexicons and NRC Emoticon and NRC Hashtags were labelled-based.

Most of the lexicons were collected from Twitter which is seen a multi-domain and can cover a large vocabulary. We collected the lexicons available publicly and categorised them into labelled and score lexicons. One approach we could have taken is to change the score lexicons into labelled ones, however, by doing so we lose the value of the scores as they can reflect strength of the polarity.

Methodology

In Figure 1, we present the methodology we used to combine the lexicons. First, we find the duplicates to each of the lexicons. In the subsections below, we explain each of the score lexicons and the labelled lexicons processes.

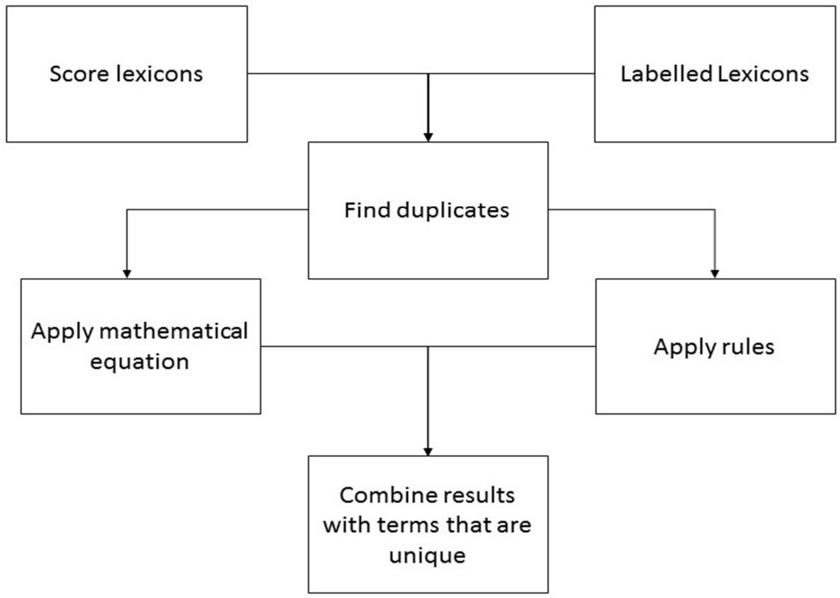

Figure 1: Methodology to Combine Lexicons

Score Lexicons

For the score lexicons, we found that a total of 189589 words were unique and 59944 words were duplicates. These duplicate terms had different scores. Therefore, to normalize the lexicons we used a mathematical formula. We first found the highest polarity score was the Arabic Hashtag Lexicon. We changed all the other lexicons to be consistent to this lexicon. The new formula made the maximum and minimum score values consistent with the Arabic Hashtag Lexicon. The mathematical formula is:

NewValue= ((OldValue-OldMin) * (NewMax - NewMin)) / (OldMax - OldMin) + NewMin

where,

- NewValue is the new value or score of the term,
- OldValue is the old value or score of the term from the lexicon we are changing,
- OldMin is the minimum score of the Arabic Hashtag Lexicon,
- NewMin is the minimum score of the lexicon we are changing,
- OldMax is the maximum score of the Arabic Hashtag Lexicon, and
- NewMax is the maximum score of the lexicon we are changing

We then averaged the scores to create a new score and combined the unique terms with the new duplicate terms. An example for our method is the word فرح farah meaning happy that occurred in all of the lexicons. As illustrated in Table 2, in AntiSenti its score was 4.86 out of the max score from that lexicon which is 7.07. Its new score became 15.78 out of 20 (the max score for the Arabic Hashtag Lexicon). We then averaged all the scores and the new score became 7.702.

Lexicon	Word	Old Score	Calculations	New Score
AntiSenti	فرح	4.866617	=((4.866617- -7.78) * (20 - -8.409)) / (7.071 - -7.78) + -	15.78271495

			8.409	
Arabic Emotion Lexicon	فرح	-0.27	=((-0.27- -5.606) * (20 - -8.409)) / (7.05- -5.606) + -8.409	3.56875158
Arabic_Hashtag_Dialetical	فرح	1.904	=((1.904- -5.877) * (20 - -8.409)) / (11 - -5.877) + -8.409	4.688732358
NRC Emoticon	فرح	0.443	=((0.443- -4.999) * (20 - -8.409)) / (5 - -4.999) + -8.409	7.052723972
NRC Hashtags	فرح	1.126	=((1.126- -6.925) * (20 - -8.409)) / (7.526- -6.925) + -8.409	7.418337831

Table 2: Example of scoring method

As seen from Table 2, the duplicated word فرح scores are all compatible and can be averaged to get a final polarity score. One of the occurrences (Arabic Emotion Lexicon) was originally a negative score but after normalising it became positive like the other lexicons.

Labelled Lexicons

The labelled lexicons consisted of 16656 unique words and 4642 duplicates. The labels included positive, negative, neutral and both (a label given when the positive is equalled to the negative). We created the following algorithm for the duplicate terms:

Frame 1. Algorithm to calculate polarity of word

```
Polarity <- " "
while there is word in file do
if (Negative > Positive AND Negative > Neutral) THEN Polarity<- "Negative"
else
if ( Neutral> Negative AND Neutral > Positive) THEN Polarity<- "Neutral"
else
if ( Positive > Neutral AND Positive > Negative) THEN Polarity<-  "Positive"
else
If ( Both EXIST) THEN Polarity <-  "Both"
else
if( Negative = Positive AND Negative > Neutral)  THEN Polarity<-  "Both"
else
if( Neutral =  Positive AND Positive > Negative) THEN Polarity<- "Positive"
else
if( Neutral = Negative AND Negative > Positive) THEN Polarity<- "Negative"
else
if(Neutral = Positive AND Neutral = Negative) ) THEN Polarity<- "Neutral"
end if
end while
```

Like previous studies we found the maximum value of the labels. If the positive, negative or the neutral were highest than the others then the word would be positive, negative or the neutral respectively.

One of our lexicons contained the label both, and if the word was labelled both in one of the duplicates then the label was always both as it held a mixture of positive and negative.

The algorithm allows us to deal with the equal occurrences in the labels. If the positives equalled the negative and were both larger than the neutral, then the words label was both. If the positive equalled the neutral and they were both higher than the negative, then the label would be positive. The reason we do this is that the positive seems to be a stronger polarity than the neutral. Likewise, If the negative equalled the neutral and they were both higher than the positive then the label would be negative. If all positive, negative and the neutral were equal, then the label is neutral.

One example to this is the word فرح farah mentioned earlier meaning happiness was positive 4 times in the lexicons and 2 times negative in the others. By using our algorithm, we can determine that the word happiness is a positive term. Another example is the word الموده elmowada which means affection. It was positive in one lexicon and negative in another. Therefore, it was labelled as both. An interesting example is the word مصير maseer meaning fate. This word occurred in three lexicons as positive, negative and neutral. The output polarity of this word in "Both" as it can be both negative and positive.

Discussion

In this study, we explored combining lexicons. With the number of lexicons available, there exist a number of overlapping and contradicting terms in sentiment lexicons. It is important to find ways to use different lexicons and to address the overlaps and contradictions. Previous studies have addressed duplicates by choosing the first instance or normalising them using a weight scale. In our research, we present a new method to combine lexicons.

Firstly, we collect several Arabic sentiment lexicons. We also chose the General Inquirer lexicon and translated it into Arabic terms. We summarized and classified the terms into scored lexicon and labelled ones. Additionally, we extract the maximum and minimum scores for each of the lexicons.

We explore a new method to merge sentiment lexicons. For the scored lexicon, we use a mathematical equation to normalize the scores. The normalization allows the maximum and minimum score to retain the highest value. Emerson and Declerck (2014) approach added weights to the polarity score rather than changing the score itself. We used a different mathematical equation than Emerson and Declerck (2014) which has yet to be tested further.

For the labelled lexicon, we create a new algorithm to combine duplicate values. Although other studies have attempted to combine the duplicates by taking the first instance or using the majority. There could exist equals. For example, when the positive lexicons are equal to the neutral. This approach allows us to deal with the equal instances.

Our final lexicons are 249532 unique terms for the score lexicon and 21298 terms for the labelled lexicon after removing the duplicates. This lexicon can be used to analyze Arabic sentiment for different domains and particularly for the Twitter data. While we have not evaluated our lexicon in this study, it can be evaluated in the future.

Although we applied this study on Arabic lexicons, it can be applied to lexicons from other languages. It can open a new era of using multiple lexicons in sentiment analysis. As stated in the literature, using a combined approach usually leads to higher results as the classifier can use the lexicon to get more accurate results. Combining different lexicons can mean that we can create a large lexicon and a sentiment analysis model that can be used for multiple domains.

One improvement that can be done to this model is use root words of the terms and analyzing it further. In Arabic, there are many Arabic words derived from the same root (Larkey et al., 2002). There exist various suffixes and prefixes. Finding the root will allow us to find more matches from the sentence. An example for this is the word فرحFarah meaning happiness. Our current lexicon has several words that are derived from it, such as فرحان Farhan meaning being happy, يفرح Yefrah and تفرح Tefrah meaning to be happy, and فرحه Farha meaning a happy event. All these words are scored separately and combining them will lead to more accurate results.

As we translated the General Inquirer lexicon into Arabic terms. The performance of this translated lexicon in sentiment analysis in comparison with other Arabic lexicons can be explored. Other lexicons can be translated and used in Arabic language.

Another improvement can be matching these dialects to the MSA words to create a more accurate lexicon. This can be done by human annotators; however, it is time consuming.

Conclusion

In this paper, we explored combining Arabic sentiment lexicons. We create a new method for scored lexicons and labelled lexicons. Firstly, we translated one of the lexicons (General Inquirer) into Arabic by human annotators. We extract the duplicates from both lexicons.

We applied a mathematical equation to the scored lexicon to normalize the data. This step allowed us to combine duplicate instances by normalizing the data according to the lexicon with the highest polarity score and changing the scores accordingly.

For the labelled lexicon, we created an algorithm to address the contradictions in duplicates. Using our method allows us to deal with equals in labelling, such as the word appearing positive and neutral the same amount of times. Other have ignored the fact that there could be equals in labelling and have only used the majority equation.

Our final lexicons are 249532 unique terms for the score lexicon and 21298 terms for the labelled lexicon. This can be used in the future to analyze Arabic tweets.

Although that the literature has stated that combing lexicons improve the sentiment analysis performance. In the future, the lexicon can be used in sentiment analysis. This method can be applied to other lexicons in different languages such as English and Spanish.

Our methodology can be improved from extracting the root words of the terms and analyzing it further. Other lexicons can also be translated and used in Arabic language to create a larger lexicon. Matching dialectical words to the MSA can allow us to create a more accurate lexicon.

Acknowledgements

This publication was made possible by the NPRP award [NPRP 7-1334-6-039 PR3] from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the author[s].

REFERENCES

Altrabsheh, N., Cocea, M. and Fallahkhair, S., 2014. Learning sentiment from students' feedback for real-time interventions in classrooms. In *Adaptive and Intelligent Systems* (pp. 40-49). Springer International Publishing.

Al-Twairesh, N., Al-Khalifa, H. and Al-Salman, A., August 7-12, 2016. AraSenTi: Large-Scale Twitter-Specific Arabic Sentiment Lexicons. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (pp. 697–705).

Balage Filho, P.P., Pardo, T.A. and Aluısio, S.M., 2013. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (STIL)* (pp. 215-219).

Cho, H., Lee, J.S. and Kim, S., 2013. Enhancing Lexicon-Based Review Classification by Merging and Revising Sentiment Dictionaries. In *IJCNLP* (pp. 463-470).

Darwish, K., Magdy, W. and Mourad, A., 2012, October. Language processing for arabic microblog retrieval. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 2427-2430). ACM.

Das, A. and Bandyopadhay, S., 2010b. Dr sentiment creates SentiWordNet (s) for Indian languages involving internet population. In *Proceedings of Indo-wordnet workshop*.

Das, A. and Bandyopadhyay, S., 2010a. SentiWordNet for Indian languages. *Asian Federation for Natural Language Processing, China*, pp.56-63

Dunphy, D.C., Bullard, C.G. and Crossing, E.E., 1974. *Validation of the General Inquirer Harvard IV Dictionary*. Harvard University Library.

El Sahar, H. and El-Beltagy, S.R., 2014, April. A fully automated approach for arabic slang lexicon extraction from microblogs. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 79-91). Springer Berlin Heidelberg.

El-Beltagy, S.R. and Ali, A., 2013, March. Open issues in the sentiment analysis of Arabic social media: A case study. In *Innovations in information technology (iit), 2013 9th international conference on* (pp. 215-220). IEEE.

El-Makky, N., Nagi, K., El-Ebshihy, A., Apady, E., Hafez, O., Mostafa, S. and Ibrahim, S., 2014, December. Sentiment analysis of colloquial Arabic tweets. In *ASE BigData/SocialInformatics/PASSAT/BioMedCom 2014 Conference, Harvard University*.

El Sahar, H. and El-Beltagy, S.R., 2014, April. A fully automated approach for arabic slang lexicon extraction from microblogs. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 79-91). Springer Berlin Heidelberg.

Emerson, G. and Declerck, T., 2014, August. SentiMerge: Combining sentiment lexicons in a Bayesian framework. In *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing*.

Kiritchenko, S., Zhu, X. and Mohammad, S.M., 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, *50*, pp.723-762.

Larkey, L.S., Ballesteros, L. and Connell, M.E., 2002, August. Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 275-282). ACM.

Liu, B., 2010. Sentiment analysis and subjectivity. In Handbook of Natural Language Processing, Second Edition (pp. 627-666). Chapman and Hall/CRC.

Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), pp.1-167.

Maurer, D.W. and High, E.C., 1980. New Words: Where do they come from and where do they go?. *American speech*, 55(3), pp.184-194.

Mohammad, S.M. and Kiritchenko, S., 2015. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, *31*(2), pp.301-326.

Mohammad, S.M., Salameh, M. and Kiritchenko, S., 2016. How Translation Alters Sentiment. *J. Artif. Intell. Res.(JAIR)*, *55*, pp.95-130.

Prabowo, R. and Thelwall, M., 2009. Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), pp.143-157.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M., 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), pp.267-307.

Wilson, T., Wiebe, J. and Hoffmann, P., 2005, October. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.