A Framework for Arabic Concept-Level Sentiment Analysis using SenticNet

Hend G. Hassan, Hitham M. Abo Bakr, Ibrahim E. Ziedan
Department of Computer and systems Engineering, Zagazig University, Egypt

Article Info	ABSTRACT
	Arabic Sentiment analysis research field has been progressing in a slow pace compared to English and other languages. In addition to that most of the contributions are based on using supervised machine learning algorithms while comparing the performance of different classifiers with different selected stylistic and syntactic features. In this paper, we presented a novel framework for using the Concept-level sentiment analysis approach which classifies text based on their semantics rather than syntactic features. Moreover, we provided a lexicon dataset of around 69 k unique concepts that covers multi-domain reviews collected from the internet. We also tested the lexicon on a test sample from the dataset it was collected from and obtained an accuracy of 70%. The lexicon has been made publicly available for scientific purposes.

Corresponding Author:

Hend G. Hassan,
Department of Computer and systems Engineering,
Zagazig University, Egypt.
Email: hend.hgh25@gmail.com

1. INTRODUCTION

Sentiment analysis (SA), a form of text classification, is the process of classifying a given document/paragraph/sentence into two or more classes. SA involves 8 key tasks [1] including subjectivity detection, where data is classified into subjective or objective data, and polarity detection where subjective data is further classified into positive, negative or mixed. Those tasks are concerned with internet users' public opinions; the data they share help gaining perspective of the overall sentiment about a specific product, service, person, etc. not only to institutions or companies but also to other internet users. However, that raw opinionated big data is unstructured and requires semantics and syntactic analysis in order to be machine understandable.

Existing sentiment analysis approaches are categorized into four main ones [2]: keyword spotting, lexical affinity, statistical methods and concept-level sentiment analysis. Keyword spotting is basically spotting keywords in the sentence and classifying it afterward. Keywords, which have positive, negative or neutral polarity, are clear sentimental words on their own like 'care', 'angry', 'glad', 'sick', etc. However, using them in a sentence may have a different sentiment other than the one they have on their own. For example ''I care for the wrong people'', care has positive polarity, but the whole sentence evokes a negative sentiment causing a misclassification error. Asides to that, the sentence may not include any keywords like ''I would never buy this book'' which indicates a negative opinion about the book yet can't be classified. Or has a misleading comparison like ''this book is as good as a hole in the head''. Briefly, this approach is known to be the most naïve one and the most popular too for its ease of implementation and accessibility.

Lexical affinity dives a little deeper in the keywords semantics than the first approach. It assigns arbitrary words with a probabilistic 'affinity' for a particular polarity. These probabilities are usually the result of training linguistic corpora. For example, 'unforgettable' might be assigned a 50% probability of being indicating a negative affect and a 12.5% probability of being indicating a positive affect and a 37.5%

probability of being indicating a neutral affect as in 'unforgettable accident' or 'unforgettable party'. As stated these probabilities are the result of training corpora; so the bigger and the more general the corpora is the more reliable and more realistic the probabilities are. This approach outperforms the keywords spotting approach for giving words realistic polarities and not just plain positive, negative or neutral.

The third approach and the most used one too to create lexicon datasets is using the statistical methods, such as the Naïve Bayes algorithm, K-Nearest Neighbor or Support vector machines(SVM). It mainly depends on training a machine learning algorithm with features like words co-occurrence frequencies, Stylistic features, etc., collected from annotated data and then test the accuracy of the algorithm used on a test sample from the same data. It is language independent thus avoids ambiguity issues associated with Arabic [3]. Yet, these methods make classification errors when tested on smaller text units such as clauses as compared with determining the polarity on the document-level [4]. Gives a quick overview of English SA research efforts from 2002 up to 2014 that are mostly made using the statistical methods. The authors also presented some of the available tools and datasets. Furthermore, [5] discussed some of the open issues in the area of SA including that there is more focus on classing the text into positive and negative only with no deeper diving in the emotions.

Concept-level sentiment analysis was first introduced by Eric Cambira to classify text based on their semantics rather than their syntactic through the use of semantic networks like ConceptNet [6] which consists of nodes representing concepts and connected with edges labeled with common sense 'taken for granted' information provided by volunteers on the internet. Cambira *et al.* [7] developed SenticNet, a semantic resource that uses common sense reasoning techniques along with an emotion categorization model and an ontology for describing human emotions to infer the polarity of different common sense concepts like 'beautiful day' or 'feel guilty'. Each concept is assigned with one float polarity value $\in$ [-1,1], followed by SenticNet2 [8] where more concepts are added allowing a deeper and more multi-faceted analysis of text while providing a four-dimensional vector (sentic vector) to each concept combined of Pleasantness, Attention, Sensitivity, and Aptitude and presented as a float value $\in$ [-1,1] along with its top-ten affectively related concepts. Then SenticNet3 [9], which contains both common and common-sense knowledge in order to boost sentiment analysis tasks such as feature spotting and polarity detection, respectively. Then SenticNet4 [10], where both verb and noun concepts are linked to primitives so that, for example, concepts such as attain-knowledge or acquire know-how or acquire-knowledge are generalized as get information. An addition that allows processing different forms of a concept that otherwise raises a not found error. The idea of using a generative word was used in other methods too. For example, [11] used synonyms lists for positive and negative words and mapped the list to one word that already has a polarity value.

To this end, we tested SenticNet4 for the task of polarity detection on a multi-domain Arabic dataset at the sentence-level and showed results outperforming other Arabic sentiment analysis works that mainly rely on other approaches. The rest of this paper is organized as follows: the next section reviews research efforts in the area of Arabic sentiment analysis; followed by a section proposes our framework to detect the polarity using SenticNet; after which a section discusses the results obtained; finally some concluding remarks and future work recommendations are presented.

2.   LITERATURE REVIEW

This section reviews the contributions to the Arabic sentiment analysis research field. Arabic, which is the formal language of over than 20 countries around the world and spoken by 300 million native speakers, is considered under-researched compared to English in the field of sentiment analysis. See Figure 1 that represents the number of Arabic/English publications per year as presented in [12], [13] and detailed in [14], [15] respectively. And Arabic is also under-resourced with respect to the amount of data on the internet knowing that Arabic has scored 4th on the number of web users after English, Chinese and Spanish ranking the highest growth rate in terms of users with 185 million of users in June 2017 according to internetwebstats's website: http://www.internetworldstats.com/stats7.htm.

Despite that recent progress, researchers focused on using the statistical methods with a special focus on supervised machine learning classifiers. They share the same methodology presented in Figure 2 while using different pre-processing and selecting different features. SAMAR [16], a system for Arabic Subjectivity and Sentiment Analysis, uses Multi-dialectal manually annotated data that covers (Maktoob chats, tweets, Wikipedia talks and web forums sentences) and does Tokenization, lemmatization and POS tagging in the Pre- processing step. Then, the system selects syntactic and stylistic features; (Unique: is set for low frequency words, Polarity Lexicon: checks the presence of positive or negative adjectives, Dialect: checks the dialects of the text, Gender: checks the gender of text whether it's male, female or unknown, User ID: checks if the author is a person or an organization and Document ID). They also made experiments with different combinations of features and the pre-processing tasks while classifying using SVM^{light} [17]

classifier. Overall results reveal improvements over the baseline performance depending on the training data. Later on, S. Ibrahim *et al*. [18] used a manually annotated data and performed normalization and stopwords removal in the pre-processing step. Then, they selected linguistic and syntactic features like term frequency, Polar word position, detecting (negation, intensifiers, questions, and supplication) terms along with using the pattern [adjective + noun] while also classifying using SVM[light] classifier.
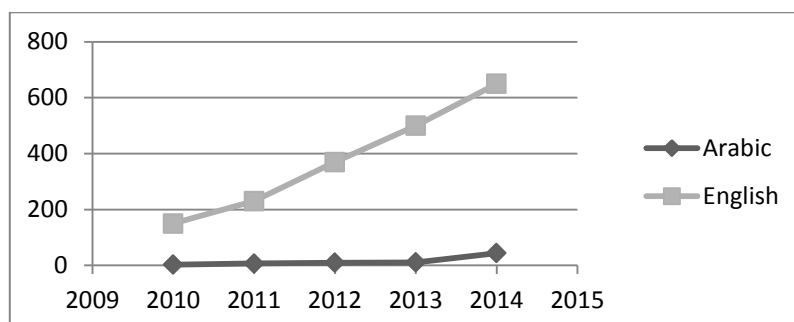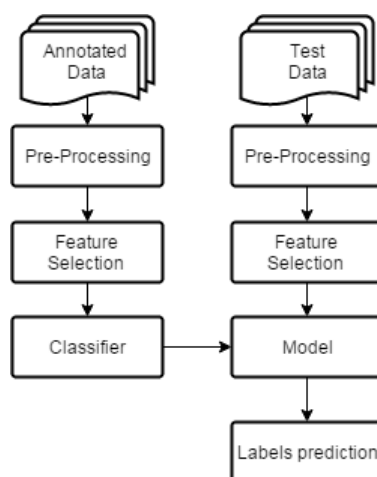

Figure 1. Number of Arabic/English publications per year

Figure 2. Supervised machine learning process

3. MODIFYING SENTICNET TO SUIT ARABIC

The framework proposed in [19] is modified to suit the task of Arabic sentences' polarity detection because Arabic natural language processing tools are trained on and made for modern standard Arabic (MSA) which is rarely used by internet users' compared to slang Arabic and other Arabic dialects. Figure 3 is an illustration of the proposed framework; Sentences are first decomposed into bi-grams then normalized and labeled with the part of speech (POS) tags. Then Syntactic patterns like [adjective + noun] are matched to extract concepts that are translated afterward into English to find a match to in SenticNet.

In order to show the effectiveness of the framework, A multi-domain public dataset is used: http://bit.ly/1wXue3C, created by ElSahar and El-Beltagy [20], covering Attraction (ATT), Hotels (HTL), Movies (MOV), Restaurants (RES#1, RES#2) and Products (PROD) reviews. We kept RES#1 as a test sample. The statistics of the dataset is presented in Table 1 showing the total number of sentences and concepts we extracted along with the number of positive, negative and mixed sentences number. Those reviews were rated by their native reviewers then were normalized into the three classes: positive, negative and mixed following the approach adopted by Pang *et al*. [21]. The main goal is to conclude the polarity of each sentence and compare it to the normalized polarity. In order to do that, sentences must be decomposed into concepts that have a match in SenticNet then the polarity of this match is read. In particular, sentences are decomposed into bigrams. If a sentence consists of only a bi-gram or a uni-gram, then it is considered a concept without further analysis. If not the concepts are extracted according to the flowchart proposed in Figure 3.
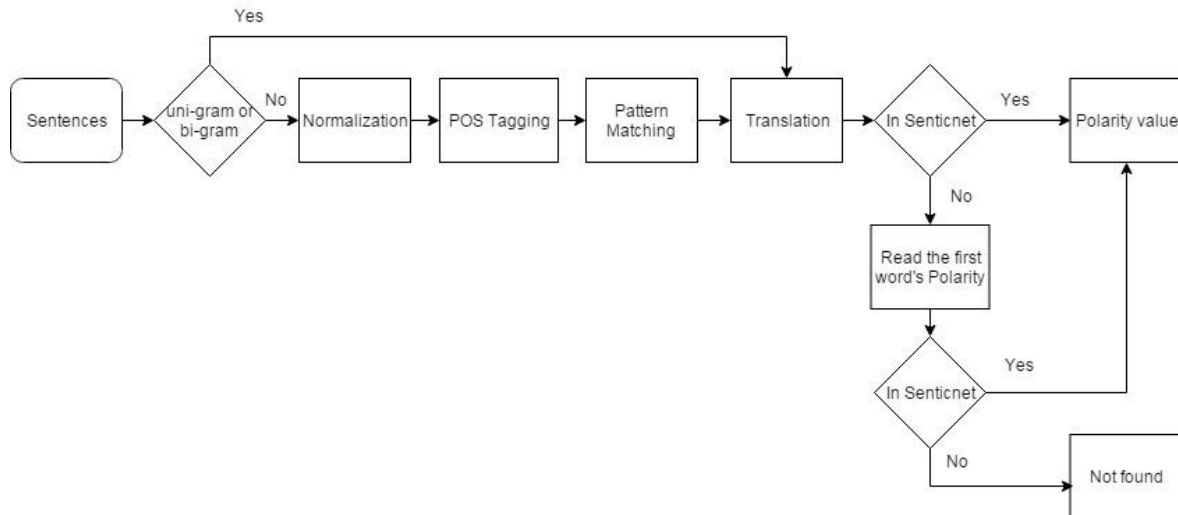
Figure 3. Flow chart of the framework to use SenticNet

Before extracting the concepts, sentences must be normalized and structured. To this end, the following pre-processing steps are followed:
1. Remove elongations.
2. Remove repetitions.
3. Remove punctuations.
4. Remove diacritics.
5. Normalize all Alef forms to ‎ا.
6. Normalize ‎ة haa to ‎ه.
7. Normalize ‎ى yaa to ‎ي.

Table 1. Dataset Statistics

	MOV	ATT	RES#2	PROD	RES#1	HTL
#Sentences	1524	2154	2642	4272	8364	15572
#Positive sentences	969	2073	2109	3101	5946	10775
#Negative sentences	384	81	268	863	2418	2647
#Mixed sentences	171	0	265	308	0	2150
#unique concepts	18511	9100	5862	4654	26000	41046

The part of speech (POS) tags of the Normalized text are then generated using MADAMIRA [22], a shallow syntactic parser that does tokenization, part of speech tagging, and base phrase chunking, and also combines some of the best aspects of MADA [23] and AMIRA [24]. By reviewing the noun phrases extracted by MADAMIRA for ATT reviews, we found that around 60% of them are unigrams 20% of which are a separation of ' ‎ال ' the Arabic definite article and about %13 are pronouns separation on the word-level like ' ‎تصميم + ه ' ('his + design'). Asides to misclassification errors, those 73% are ineffective as concepts. Thus we used a hand crafted syntactic patterns following the work of ElSahar and El-Beltagy [25] that extracts slang terms (words/expressions) and transliterated English written in Arabic letters like ' ‎أوفر ' that is transliterated from 'over'. Their work depends on creating a set of lexico-syntactic patterns by using standard tags like Negator [Neg], person reference [PR], Personal Pronoun [PP], Demonstrative Pronoun [DP],

Intensifier [Ints], Conjunction [Conj], Strong subjective [SS] and the extracted Subjective Expression is {SE}. For example, "Respectable and very {polite}" would match the pattern: [[SS] [Conj] {SE} [Ints]], having 'polite' as the extracted term. They created 11 different patterns with a finite set of terms in each tag and were able to extract 633 unique terms out of 7.5M twitter corpus. In order to be able to match more patterns, we added to those tags the part of speech tags labeled by MADAMIRA comprising different patterns as detailed in Table 2. For example [Adjective + [Ints]] would match 'very excellent'.

We also benefited from the fact that the Arabic language has embedded ' ‎ال ' in definite nouns and two consecutive nouns are usually a concept like ' ‎السنوات الاخيرة ' ('recent years') that can be extracted easily using the pattern [{} ‎ال {} ‎ال]. Although using syntactic patterns is considered a heuristic method, it extracted

better concepts than using MADAMIRA's noun phrases and verb phrases alone considering the ambiguities associated with Arabic.

Table 2. Set of Patterns with Matched Examples

Pattern number	Syntactic pattern	Pattern's Exception	Example of a match	Bing Translation	Concept
			Noun phrases		
P1	الـ { } الـ { } الـ وال { } الـ { } وال للا { } الـ { } للا كال { } الـ { } كال فال { } الـ { } فال بال { } الـ { } بال لل { } الـ { } لل	Second word in [التى ,إلذى, إللى, etc.]	السنوات الأخيرة والوسواس القهري للأفكار الجديدة كالقوي الكونية فالمسئول الحكومي بالدرجة الأولى للوجبات السريعة	recent years and obsessive compulsive disorder to new ideas as cosmic powers the government official in first class for fast food	recent_year obsessive_compulsive_dis new_idea cosmic_power government_official first_class fast_food
P2	الـ { } Adjective وال { } Adjective فال { } Adjective		الفيلم جديد والسعر عالي فالفيلم جيد	New movie And the price is high It is a good movie	new_movie price_high good_movie
P3	ين { } ين { } ين	[بين ,إلذين ما, بين etc.]	بمقاسين مختلفين	With different sizes	different_size
P4	ون { } ون { } ون		العاملون مدربون	Trained workers	Trained
P5	ان { } ان { } ان	[وان ,كان, إن, etc.]	شخصيتان فريدتان	Unique personalities	Unique
P6	Adjective + [Ints]		ممتاز جداً	Very excellent	Excellent
P7	Noun + adjective		بحركات سريعة	Quick movements	quick_movement
P8	[P_pron] + Adj. [P_ref] + Adj. [D_pron] + Adj.	Pattern followed by 'من'	انت محظوظ ناس معينة هو الأفضل	Lucky you Certain people Is best	Lucky certain best
			Verb Phrases		
P9	Verb + noun		وقضيت وقت	I spent time	spend_time

```
Procedure:  polarity detection
Input:
English translation of the patterns
Output:
Polarity, Pleasantness, Attention, Sensitivity, Aptitude, Semantics
Begin:
    For each pattern in the sentence:
        Remove punctuations
        Remove stopwords
        lemmatize the nouns
        If the first word's tag is verb
            Lemmatize the verb
        Search for a match in SenticNet
    End
    Sum the polarities of each sentence
    If classifying into three classes
        If (sum >= 1)
            Positive
        Else if (sum <=- 1)
            Negative
        Else
            Mixed
        End
    Else If classifying into two classes
        If (sum >= 0)
            Positive
        Else if (sum <0)
            Negative
        End
    End
End
```

Figure 4. Pseudo code for polarity detection

Next, we used Microsoft Bing translator to translate the matches to English. Having English on the output side of the machine translation system and not translating concepts from SenticNet into Arabic avoids the ambiguity of different dialect candidates and different sentence structures; Arabic is one of the languages

that has multiple sentence forms [subject-verb-object (SVO), verb-subject-object (VSO), verb-object-subject (VOS) asides to the possibility of having a correct sentence dropping a verb or copula].

At last, the translated extracted concepts follow the steps presented in Figure 4 to match the same form of SenticNet's concepts where nouns are singular and verbs are lemmatized. If a match is found in SenticNet then the polarity value is read. If not a search for a match for the first word of the concept is done as it is usually an adjective for example 'فندق رائع' (wonderful hotel) would extract 'wonderful' if the whole concept 'wonderful_hotel' is not found in SenticNet.

4.   RESULTS AND DISCUSSION

In order to properly evaluate the performance of the proposed framework, we used the leading measuring methods in the NLP Classification process: precision, recall, f-measure and accuracy that are shown by values in table3 respectively for each dataset in case of 2-class classification problem (positive and negative) and 3-class classification problem (positive, negative and mixed).We highlighted the datasets with the best scores revealing that the 2-class classification problem has better results than the 3-class classification problem for the same dataset. The same result was also obtained in [20]. And to show the difference between our method and existing ones, we compared these results with the ones obtained in [20] in which they used the same dataset we used but while using the statistical methods (See Table 4). The reported average accuracy in Table 3 is the average of all accuracies reported after using different lexicon based features. Their reported accuracy is a result of training 80% of the data with a machine learning classifier and calculating accuracy on a 20% test sample from the same data with the classifier. Furthermore, best accuracy score in the 2-class classification happens for the ATT dataset. This could be explained by the fact that it has more concepts extracted as compared to RES#2 that has more sentences but fewer concepts and for which it scored the second best accuracy.

Table 3. The Value of Precision, Recall, F-measure and accuracy respectively for each Dataset

		P	R	F1	Acc.	ElSahar's work average accuracy
3-class	ATT	There are no mixed polarity reviews in the ATT dataset				Not mentioned
	PROD	.57	.45	.49	.45	0.51
	MOV	.51	.62	.54	.62	0.47
	RES#2	.72	.72	.71	.72	0.57
	HTL	.55	.62	.58	.62	0.64
2-class	ATT	.96	.86	.91	.89	Not mentioned
	PROD	.78	.91	.84	.73	0.74
	MOV	.73	.91	.81	.70	0.69
	RES#2	.91	.91	.91	.85	0.81
	HTL	.81	.87	.84	.73	0.85

Table 4. Comparison between our Results and ElSahar's Results

	Average accuracy		Dataset lexicon
	3-class	2-class	
ElSahar's work	.56	.77	2k un normalized entries
Proposed framework	.60	.75	96k unique entries

Figure 5 is a boxplot of the number of words and concepts for each dataset and it shows that ATT has more words in the sentence than RES#2 causing more concepts. Although MOV dataset has relatively more concepts, it scored last. That can be explained by the fact that it has the longest review length as it has 2530 words in one of the reviews. The 3-class classification problem has the same ranking order except for the PROD dataset that has fallen behind as it has uni-gram reviews.

On the other hand, ElSahar 's lexicon [20] has around 2000 entries (uni-grams and bi-grams) that are not normalized nor lemmatized;'أنصح' (I recommend) , 'انصح بشراء' (I recommend to buy) , 'انصح به ' (recommend it) ,' انصح بها ' (I recommend it/female pronoun),' انصحكم ' (I recommend you) are all entries and all of them has the same lemma 'recommend' while we were able to extract around 69 k unique entries after removing redundancy from the different datasets. Furthermore, we used a test sample from the dataset (RES#1) in order to validate our lexicon by following the same steps in the framework while skipping the translation step as shown in Figure 6. We were able to match 68% of the concepts extracted from RES#1 in the lexicon. The accuracy obtained was 70% and the precision was 70% with a recall of 100% and an F-measure of 82%.
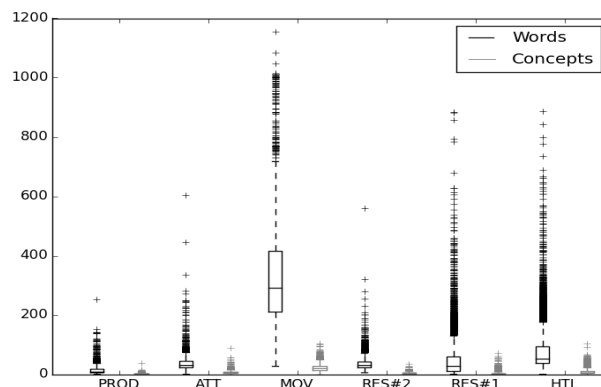
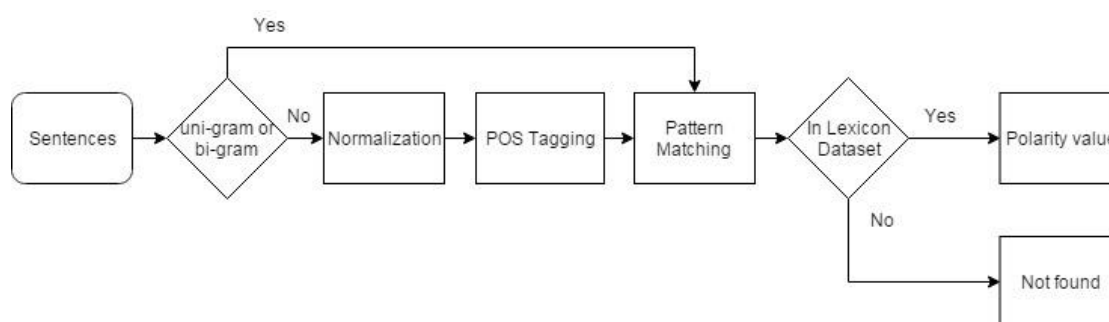Figure 5. Boxplot's section showing the number of words and concepts for each dataset

Figure 6. Flow chart for the testing framework of the lexicon dataset

5. CONCLUSION AND FUTURE WORK

A novel framework for concept-level sentiment analysis was introduced to detect the polarity of Arabic sentences using Senticnet. The framework is created so that it can handle ambiguity issues associated with Arabic including the fact that slang Arabic lacks syntactic rules and tools to deal with and it also doesn't include using any machine learning algorithm.The framework was tested on a multi-domain dataset covering public reviews scrapped from the internet. The results showed promising performance as the accuracy reached 89%. it also outperformed other research works in terms of detecting the polarity of a sentence without having prior annotated data. In the future, we plan on Handling Polarity inversion terms such as negations and also defining the scope of each negating term along the sentence.

REFERENCES

[1] E. Cambira, *et al.,* "The CLSA Model: A Novel Framework for Concept-Level Sentiment Analysis", *16th international Conference,CICLing 2015,* Part II, pp. 3-22, 2015.

[2] E. Cambira, "An Introduction to Concept-Level Sentiment Analysis", *The 12th Mexican International Conference on Artificial Intelligence*, Part II, pp. 478-483, 2013.

[3] N. Y. Habash, "Introduction to Arabic Natural Language Processing", *Synthesis Lectures on Human Language Technologies 3.1,* 2010.

[4] A. Dhokrat, *et al.*, "Review on Opinion Mining for Fully Fledged System", *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 4, no. 2, pp. 141-148, 2016.

[5] P. Kumar and S. Nandagopalan, "Insights to Problems, Research Trend and Progress in Techniques of Sentiment Analysis", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 5, pp. 2818-2822, 2017.

[6] H. Liu and P. Singh, "ConceptNet — A Practical Common Sense Reasoning Tool-kit", *BT Technology Journal*, vol. 22, no. 4, pp. 211-226, 2004.

[7] E. Cambria, *et al.,* "SenticNet: A Publicly Available Semantic Resource for Opinion Mining", *AAAI Fall Symposium: Commonsense Knowledge*, vol. 10, 2010.

[8] E. Cambria, *et al.,* "SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis", *FLAIRS conference*, pp. 202-207, 2012.

[9] E. Cambria, *et al.,* "SenticNet 3: A Common and Common-Sense Knowledge Base for Cognition-Driven Sentiment Analysis", *Twenty-eighth AAAI conference on artificial intelligence*, 2014.

[10] E. Cambria, *et al.,* "SenticNet 4: A Semantic Resource for Sentiment Analysis based on Conceptual Primitives", *COLING,* pp. 2666-2677, 2016.

[11] P. Arora, *et al.,* "An Approach for Big Data to Evolve the Auspicious Information from Cross-Domains", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 2,pp. 967-974, 2017.

[12] A. AlOwisheq, *et. al.*, "Arabic Sentiment Analysis Resources: a Survey", *Social Computing and Social Media: 8th International Conference*, pp. 267-278, 2016.

[13] K. Dashtipour, *et al.,* "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques Cognitive Computation", vol. 8, no. 4, pp. 757-771, 2016.

[14] Al-Twairesh, *et al.,* "Subjectivity and Sentiment Analysis of Arabic: Trends and Challenges", *11th International Conference on Computer Systems and Applications (AICCSA)*, IEEE/ACS, pp. 148-155, 2014.

[15] K. Ravi and V. Ravi, "A Survey on opinion mining and sentiment Analysis: Tasks, Approaches and Applications", *Knowledge-Based Systems*, vol. 89, pp. 14-46, 2015.

[16] M. A. Mageed, *et al.,* "SAMAR: A System for Subjectivity and Sentiment Analysis of Arabic Social Media", *Computer Speech & Language*, vol/issue: 28(1), pp. 20-37, 2014.

[17] T. Joachims, "Svmlight: Support vector machine", Cornell University, 2008.

[18] H. Ibrahim, *et al.*, "Sentiment Analysis for Modern Standard Arabic and Colloquial", *International Journal on Natural Language Computing (IJNLC)*, vol. 4, no. 2, pp. 967-974, 2015.

[19] E. Cambria and A. Hussain, "Sentic Computing A Common-Sense-Based Framework for Concept-Level Sentiment Analysis", *Springer*, vol. 1, 2015.

[20] H. ElSahar and S. R. El-Beltagy, "Building Large Arabic Multi-domain Resources for Sentiment Analysis", *16th international Conference,CICLing 2015*, Part II, pp. 23-34, 2015.

[21] B. Pang, *et al.*, "Thumbs up? Sentiment Classification using Machine Learning Techniques", *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing,* vol. 10, pp. 79-86, 2002.

[22] A. Pasha, *et. al.*, "MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic", *LREC*, 2014.

[23] N. Habash, *et. al.*, "MADA+TOKAN:A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization", *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools(MEDAR)*, Cairo, Egypt, pp. 102-10.

[24] M. Diab, "Second Generation AMIRA Tools for Arabic Processing Fast and Robust Tokenization, POS tagging, and Base Phrase Chunking", *Proceedings of KONVENS*, pp. 39-52, 2012.

[25] H. ElSahar and S. R. El-Beltagy, "A Fully Automated Approach for Arabic Slang Lexicon Extraction from Microblogs", *15th International Conference, CICLing 2014*, Part I, pp. 79-91, 2014.