International Journal of Engineering & Technology, 7 (3.13) (2018) 128-132

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Review on Sentiment Analysis in Arabic Using Document Level

Ibrahim Awajan [1]*, Mumtazimah Mohamad [2]

[1&2] Faculty Informatic &Computing, Universiti Sultan Zainal Abidin, Besut Campus, Terengganu, Malaysia
*Corresponding author E-mail:*si2150@putra.unisza.edu.my

Abstract

With the rapid increase of web contents and the spread of social media and microblogs such as Twitter and etc.., the focus on the sentiment analysis (SA) is deeply being studied. The research on Arabic sentiment analysis is progressing very slow in compared to English sentiment analysis. The former has recently attracted a considerable concentration of researchers. In this respect, this paper aims to present a brief review of some major works that have addressed the document-level sentiment analysis in Arabic. This review includes research and studies published during the period of 2011-2017 for different approaches of document-level sentiment analysis.

Keywords: *Classification; Information Extraction; Opinions; Polarity; Sentiment Mining.*

1. Introduction

With internet availability and exponential increase in it is usage, the web has become a platform to read and write information. This textual information is growing every day on the internet, so there is an abundance of web forums, social media and personal blogs and etc.

The users of the internet are no longer just consumers of web content, but also producers of it. Their contributions become very important to enrich the web content. These users express their opinion and share information on different fields and topics.

Due to the increasing of the opinions, reviews, feedbacks and emotions on the web, exploring and analyzing these information becomes very important in determining the best decision making by other users ranging from products, movies, education, health and politics to hotels and different services that help people, companies, institutions, and states making their decisions or prediction. According to Pew surveys (2014) [1], around 81% of adult American internet users use the internet to browse information about products or services that they want to buy. This information can be used to create a relationship between the products that are purchased [24].

Therefore, automating the process of text sentiment analysis has become useful. People will be able to access the opinions and sentiments about a specific topic in a good manner, rather than reading the reviews to obtain the final opinion. This new research domain is called sentiment analysis or opinion mining. Sentiment Analysis(SA),extracts, analyzes and evaluates language and its inflections to determine attitudes, opinions and emotions for reviews or documents that related to a business, product , service or any topic, by including many fields such as statistics, Natural Language Processing (NLP) and machine learning to determine if an expression is positive, negative, or neutral, and to determine its degree, so that we can benefit from the results and employ it in the field or domain that we interested in. [2] Proved that sentiment analysis useful for many Natural Language Processing (NLP).

In sentiment analysis initially, the content of the text is determined whether it is subjective or objective [2]. This step is called subjectivity classification. Then the second step is to analyze the subjective text and its purpose to determine which of the sentiment polarities it has [2]. In the sentiment classification the type of data used varies from one domain to another., So, if a sentiment analysis system for customer reviews works well may not work with movie reviews, and also from one language to another [3]. Our paper would be focus on Arabic language where it is considered as one of the sematic languages, and as a mother tongue of about 22 countries and used by more than 325 million persons in daily communication [4]. The Arabic script also the second most familiar script in the world after Latin [5]. Further, Arabic language is one of the fastest growing languages on the internet, it is one of the top ten languages on the internet (Figure 1). These statistics were captured in 2017 according to the Internet World State (2017).

Nevertheless, in the area of sentiment analysis, there are many research study deal with English language, but the Arabic sentiment analysis is still in its early stages and there are a few researches have been conducted in this direction. Although it started after 2008, the studies on sentiment analysis in Arabic have increased over the last five years. This may be due to relatively limited research funding and scholarly work in this area compared to English language studies. In addition to the morphological complexities of the Arabic language, since the Arabic language can be divided into three main types are Classical Arabic (CA), Modern Standard Arabic MSA, and Dialect Arabic (DA) [8].CA is the language of the Qur'an the oldest version of Arabic. The MSA is the current formal Arabic language of Arabic countries used in media, education, newspapers, and books.

DA is the colloquial language and there is a difference in the Dialect Arabic from one region to another in the Arab countries. There are similarities and differences between MSA and CA since MSA has the same syntax and morphology of CA [9].

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

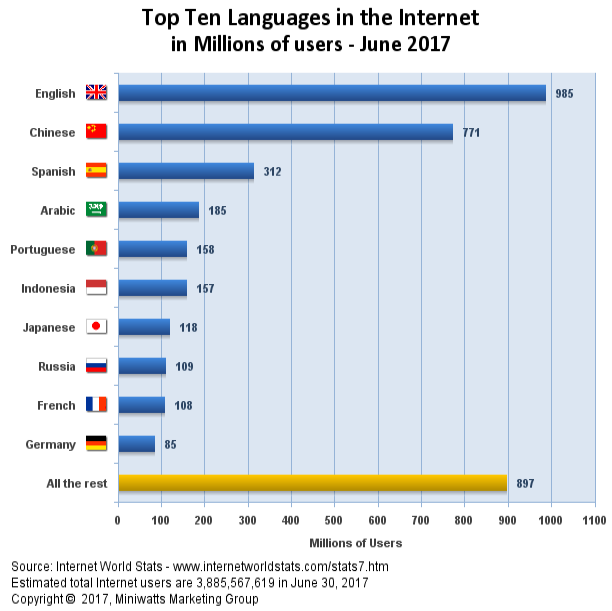


Fig. 1: Top ten languages on the internet (Internet World Stats - www.internetworldstats.com/stats7.htm)

The words of DA mostly derived from MSA. Recently, some people have become writing MSA as well as Arabic dialects using Latin Characters which is widely used in social media platforms. It is called Arabizi or Romanized Arabic.

From the foregoing, it is clear that we face challenges and problems with sentiment analysis for the Arabic language, especially for dialectal varieties and morphological complexities of the Arabic language, where these challenges are compounded by the presence of orthographical variants and multiple dialects, which are often used along with the formal written language.

In sentiment analysis, the classification process can be applied at three levels of the text (sentence, document, and aspect). Our work would be concentrated on the document level.

The rest of this paper is structured as follows: Section II presents a literature review of related studies, section III, contains material and methodology of document-level sentiment analysis, in Section IV, the previous study results are discussed. We finish by a conclusion in the last section.

2. Literature Review

Many papers have presented different approaches for Arabic Sentiment Analysis where they dealt with the problem of sentiment classification in different ways. In general, the sentiment Analysis can be performed at three different levels of the text:

i. Sentence-Level Sentiment Analysis: In this level, each sentence may have a different opinion, so the polarity will calculate for each sentence where each sentence is considered as a separate unit.
ii. Aspect-Level Sentiment Analysis: Sentiment and opinions of individual phrases and sentences are determined about a particular aspect, where product features are identified and extracted from the source data.
iii. Document-level sentiment analysis: it is the most common level. As clear from it is named it assumed that the author of the document has an opinion on one main object expressed in the entire document. where express the negative opinion in negative value and positive opinion in the positive value. So at this level, we can go beyond the problem of determining the boundaries of sentences that facing the previous level. But the document level has special challenges, most notably is containing the article for more than one opinion, also the inverse sentiments in the same article, where the opposite opinion may invalidate the main opinion like (It could have been a good movie but the acting is well done...etc.).

In sentiment analysis, there are three approaches that commonly has been used to perform sentiment classification.

First, Machine learning approach. Machine learning is a technique that gained the interest of researchers because of its accuracy and adaptability. where often depend on supervised classification approaches. It uses linguistic features and applies the famous ML algorithms. Initially, known data rules are used to train the algorithm then applied this algorithm on an actual dataset. on the other hand, Lexicon-based Approach also has commonly used by others. is an unsupervised technique.

Authors of [13] performed a binary sentiment classification utilizing three proposed classifiers: SVM, NB, and KNN. Before the classification step. They applied preprocessing phase by proposing three methods are Remove Similar, Remove Farthest and Remove by Clustering. Where the algorithm of each method contains four different steps. Authors used two corpora, the first one was ACOM (Arabic Corpus for Opinion Mining) that includes two Arabic datasets for two different domains, the second one was SINAI corpus which developed by Rushdi-Saleh et al [12]. That are different in domain, document size, unbalance rate and language.

The authors mentioned that proposed preprocessing improved the classification performance. The results of their experiments showed that k-NN is sensitive to unbalanced datasets but SVM is highly sensitive. On the contrary, NB is not sensitive to unbalanced data sets.

Authors of [15] performed sentiment classification by two forms sentiment, polarity classification, and rating classification. They applied machine learning using SVM, MNB, and BNB. They also used simple features. Authors presented large Arabic sentiment analysis dataset by collecting dataset from www.good- reads.com and created a Large-scale Arabic Book Review (LABR) that contains over 63,000 book reviews.

Notwithstanding they utilized simple classifiers and features, sentiment polarity classification achieved (90% accuracy), but for rating classification, there was much room for improvement in rating classification (50% accuracy).

Authors of [17] proposed machine learning approach by using (SVM, KNN, BNB, MBN and stochastic gradient descent). Authors collected dataset from twitter, that contains 10,006 Arabic Tweets. where they applied the classifiers on the balanced and unbalanced datasets and performed two set experiments: first one, four-way sentiment classifications which are classified texts as objective, subjective positive, subjective negative, and subjective mixed and the second one, Two stage classification. where SVM classifier was the best one among all classifiers based on results of experiments. Even so, their experiments did not show good results by using n-grams features in multi-way classification.

Authors of [19] applied machine learning approaches to classifying tweets using Naive Bayes (NB) and Support Vector Machines (SVM) with some selection schemes like TF-IDF and BTO. Authors applied Preprocessing (tokenization, removal stop word, light stem and filter token by length) on the dataset. They used the polarity to classify tweets (positive, negative and neutral) where they collected the used tweets from public timelines.

The results of their experiments showed the effect of applying text Pre-processing in sentiment analysis classification as a key factor. Also, that SVM classifier has promising results.

On the other hand, Lexicon-based Approach has commonly used by others. It is an unsupervised technique so no need training data and alternatively depend on semantic orientation. where It utilizes a list of predefined words, each word is related to a specific sentiment, Thus the whole process becomes faster. This Approach divided into corpus-based and dictionary-based to analyze the sentiment polarity

Authors of [14] proposed A lexicon-based approach to perform subjectivity classification for both Modern Standard Arabic (MSA) news articles and dialectal Arabic microblogs from Twitter. Since they used lexicons (MPQA) that is an existing English subjectivity lexicon, and a manually created Arabic (ArSenti) because the MPQA lexicon had many translation errors. the authors used

stemming and POS tagging. and applied positive and negative sentences for polarity classification experiments. The results of their experiments showed an improvement in sentiment classification and subjectivity for Arabic tweets.

Authors of [16] proposed a lexicon-based approach to sentiment classification of Arabic texts, particularly for Egyptian texts. This approach contains three steps: building a sentiment lexicon, assigning weights to words and determining semantic orientation.

Authors used two datasets, the first dataset (Twitter dataset) contains 500 tweets and the second dataset (Dostour dataset) contains 100 web comments. These datasets used to evaluate two unsupervised classification algorithms the Sum method and the Double Polarity (DP) method. Their experiments gave good results (83.8% accuracy) on a Twitter dataset.

The evolution of sentiment analysis has shown researchers exploring the possibility of using a hybrid approach. this approach is suitable where is a combination of machine learning and lexicon based approaches. This approach has the potential to combine the speed of lexical based approach and performance of a machine learning approach.

Authors of [11] proposed combined approach contains lexicon-based method, maximum entropy (ME) method, and k- nearest neighbor (KNN) method. Since they used these methods combined. This proposed approach has succeeded in producing better results. where they utilized the lexicon-based method to classify the largest number of documents as possible. Then the maximum entropy method used the resulting classified documents from the lexicon-based method as a training set to classify some other documents. At last, k-nearest neighbor method utilized the resulting classified documents from the lexicon-based method and ME as a training set to classify the remaining documents.

The authors have tested an experiment involved their different domain dataset which are education, politics, and sports which is totally 8793 Arabic statements from 1143 post. Generally, with combined method, the result showed better improvement for 80% accuracy compared to the use of single or lesser combination.

Authors of [18] applied a semi-supervised approach to building a multi-domain lexicon of entries which covered four multi-domains reviews datasets (hotel reviews, movie reviews, restaurant reviews and product reviews). The results have shown that the SVM was high performing classifier. While KNN was the worst, but the combination of the lexicon based features with the other features gave best performing feature on the total accuracy, with 2% in case of Counts and 10% in cases of TFIDF and Delta-TFIDF. Also, Authors pointed out that documents rich in polar expressions improve the performance of the algorithm to give better results.

3. Techniques and Methodology

In this section, we aim to understand the methodology of sentiment analysis and present the sentiment analysis process, also we will present the applied techniques and in document-level sentiment analysis.

We will start with the general sentiment analysis steps as shown in figure 2.

The first step includes collection of data that may involve any format of data (Word, PDF, XML, HTML, etc.) that we need to process.

These data are converted to text and then pre-processed. Pre-processing step is very important in sentiment analysis. In order to obtain better results, at this step the raw data is cleaned and prepared by reducing its complexity as follows to prepare the final data for the classification step:

- **Extraction (Tokenization):** This method aims to tokenize the document content into individual term(word)Segmentation of sentences.
- **Normalization:** In the Arabic language, there are some characters written in different ways such as (Alef Hamza below "ٕ" and Alef Hamza above "ٔ"). So, aims this step to normalize the spelling of these characters.
- **Non-Arabic removal:** Ignore data that contain letters {A-Z, a-z}.
- **Remove stop words:** Arabic stop words are not useful in the search. So, they are removed. Where removing stop words saves time and space.
- **Negation handling:** Convert positive sentiment to negative, or from negative to positive by using special words.
- Expand acronyms: Expand acronyms to their original words by acronym dictionary.
- **Stemming:** Root-based stemming, Light Stemming, Statistical stemmer and a hybrid approach.
- **Filtering:** By removing URL links, special words in Twitter, user names in twitter, Repeated lettered and emoticons.

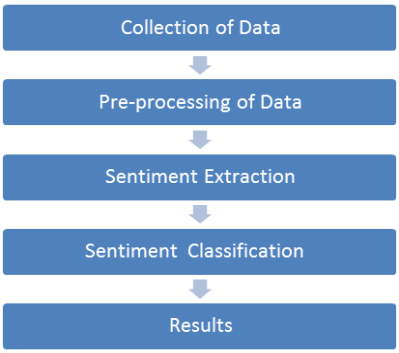


Fig. 2: A general sentiment analysis process

The classification step can be done at three levels of the text:

1. **Sentence-Level Sentiment Analysis:** Based on the classification of sentiment and opinions expressed in each sentence.
2. **Aspect-Level Sentiment Analysis:** Aims to determine the sentiment and opinions of individual phrases and sentences about a particular aspect.
3. **Document-level sentiment analysis:** this level assumes that the author of the document has one expressed opinion in the entire document on the main subject. Document-level sentiment analysis divided into supervised approach, unsupervised approach and semi-supervised approach:

A. **Supervised learning:** In this approach, there are a finite set of classes into which the document should be classified and training the classifier on the text polarity prediction [21]. Since the classifier will learn from the training data and (sometimes) builds a model which is later used to classify the documents of the testing set, the system learns a classification model by using one of the common classification algorithms such as:

- Decision tree classifiers: (D-Tree).
- Linear classifiers: SVM (Support vector machines), NN (Neural Network).
- Rule-based classifiers.
- Probabilistic classifiers: NB (Naive Bayes), BN (Bayesian Network), ME (Maximum Entropy).

This classification is then used to tag new documents into their various sentiment classes. When a numeric value (in some finite range) is to be assigned to the document then regression can be used to predict this value (like, the Amazon stars ranking system).

B. **Unsupervised learning (lexicon-based):** Lexicon-based is divided to dictionary-based approach and corpus-based approach, which use statistical or semantic methods to find sentiment polarity, that can build either manually or automatically [22].

Unsupervised learning for document-level Sentiment Analysis based on semantic orientation (SO) of phrases within the document. Associated every word with a polarity value (positive or negative). There are two weighting algorithms to assign a sentiment weight:

Sum method: It is done by extracting the sentiment polarity of each word in the document where the weight of 1 will assign for positive words while the weight of -1 will assign for negative ones. The total polarity of the document is calculated by accumulating the polarity score of each word in the document [23].
Double Polarity (DP) method: For each word in the lexicon will assign both a negative and a positive weight. For example, if a negative word in the lexicon has a weight of -0.7, then it will be: (1+(-0.7)=0.3) the positive weight. As well, a positive word of 0.9 weight has a -0.1 negative weight. The polarity of the document is calculated by gathering all the negative weights and all the positive weights in the document. Then the final polarity is calculated depending on the greater absolute value of the combination result.

C. **Semi-supervised approach:** This approach combines Supervised learning and Unsupervised learning methods.

Finally, the results: The goal of sentiment analysis is to extract meaningful information from unstructured document text. The results of text can be displayed after the analysis is completed in a graphical form like a bar chart, pie chart, and line graphs. Also can be analyzed the time and graphically presented.

4. Discussing

Table 1: Research on Document-level sentiment analysis for the Arabic language

N	paper	year	Dataset	The type of classification	Approach	Accuracy
1	11	2011	Multi domain reviews	Based Lexicon, ME and KNN	Semisupervised	80%
2	13	2012	Aljazeera	SVM, NB, KNN	Supervised	74%
3	14	2013	Twitter	NB	Supervised	80%
4	15	2013	Book reviews	SVM, MBN, BNB	Supervised	91%
5	16	2013	Twitter	Based Lexicon	Unsupervised	83%
6	17	2015	Twitter	SVM, MNB, BNB, KNN, passive aggression, SGD, LR	Supervised	69%
7	18	2015	Multi domain reviews	Linear SVM, LR, Bernoulli Naive Bayes, KNN , stochastic gradient descent.	Semisupervised	88%
8	19	2016	Twitter	SVM, NB	Supervised	89%
9	20	2017	newspaper	NN, NB, k-NN, SVM, RF, CT, LR, and CN2 Rules	Supervised	84%

Sentiment Classification at document level has the largest share of research studies by using either supervised learning, unsupervised learning, or semi-supervised learning.

- **Supervised Learning:** Most research on sentiment analysis using supervised learning algorithms to create sentiment classification models that are require labeled data. As indicated in table [1], The supervised learning approach has been applied repeatedly. Although, labelled data are often difficult and costly to acquire. And often used SVM and NB as classifiers. Also, the n-grams have been used as a feature repeatedly. The results have shown the effectiveness of the supervised learning approach in most research. However, this study shows that obviously the less accurate was for paper [17] (69.10%). Authors used four-way sentiment classification task in four classes (objective, subjective positive, subjective negative, and subjective mixed). Since using this 4-way consider as advantage and more challenge than 3-way classification task with classes of sentiment polarity (positive, negative and neutral). On the other hand, this paper applied machine learning algorithms (SVM, MBN, BNB, KNN and stochastic gradient descent) on the balanced and unbalanced datasets. However, It gave bad results, Since used n-grams as unique features in multi-way classification. But the same algorithms gave better result in other research. It seems that using n-grams as unique features in multi-way classification is the reason to give bad results. We would like to mention that in this paper, SVM is considered as the best classifier, since it gave accuracy (69.1%).
- **Unsupervised Learning**: This approach utilizes mainly for classifying the unlabeled data. It is used because the labeled data that used in supervised approach is more difficult than collecting data for training. But we have a challenge with the sentiment dictionaries for Arabic language. As we already mentioned, there is no abundance of studies and research is conducted in Arabic language, Unlike English. So English dictionaries are often used to create Arabic dictionaries by using machine translation. However, It does not give the desired results. That is why the future studies must work to improve Lexicon-dictionaries for Arabic language and expand it to include the missing terms. The studies demonstrate this approach does not give better performance than supervised learning approach. As indicated in table [1], Paper [16] gave a performance with accuracy (83%), especially on twitter datasets.
- **Semi-Supervised Learning:** This approach exploits the advantages of both labeled and unlabeled data during the training stage. This approach creates cooperation between unsupervised and supervised learning by adjusting for the shortfall of labeled cases with unlabeled ones, So Semi-Supervised Learning is good to build sentiment classification models. Therefore, we expect that shedding light on this approach in the future could lead to promising results. As indicated in the table [1], Studies [18] and [11], Authors proposed a combined approach that combines lexicon and supervised learning methods. Their experiments showed that this proposed approach gave better results than using each method alone.

5. Conclusion

In this paper, we have surveyed about the important document-level sentiment analysis studies in the Arabic language. Where Arabic sentiment analysis is still in its early stages but the studies on sentiment analysis in Arabic have increased over the last five years. However, there are challenges with the morphological complexities of the Arabic language compared to the English language. We concentrated on document-level sentiment analysis studies. It is the most common level where it assumed that the author of the document has an opinion on one main object expressed in the entire document either negative or positive. So the document level has special challenges, such as containing the article for more than one opinion, also the inverse sentiments in the same article.
Also, we presented sentiment analysis process and applied techniques in document-level sentiment analysis, where studies have shown the importance and effect of preprocessing on sentiment classification. For sentiment classification, there are three types of approaches (Supervised, Unsupervised and Semisupervised) Among these approaches the supervised learning is the most used where gained the interest of researchers because of its accuracy and adaptability where SVM classifier was most applied. On the other hand, unsupervised learning has used by others because it is no need training data and alternatively depends on semantic orientation. Semisupervised learning has the potential to combine the speed of unsupervised learning and performance of supervised learning.
Concerning the results of studies, their findings were interesting. But as we mentioned earlier there are many challenges in the nature of the Arabic language that must be taken into consideration to get better results in future research.

References

[1] Purcell, Kristen, and Lee Rainie. "Americans feel better informed thanks to the Internet." Pew Internet Research Project (2014).

[2] Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, volume 2(1-2): pages 1{135, Jan. 2008. ISSN 1554-0669.

A. Aue and M. Gamon.Customizing sentiment classi_ers to new domains: A case study. In Proceedings of Recent Advances in Natural Language Processing (RANLP), volume 1, pages 207{218. Citeseer, 2005.

[3] J. Stokes and A. Gorman, "Encyclopedia of the peoples of africa and the middle east, 2010," OnlineEdition, The Safavid and Qajar dynasties, rulers in Iran from, vol. 1501, p. 707.

[4] L. McLoughlin, Colloquial Arabic (Levantine). Routledge, 2009.

[5] Sarah O. Alhumoud, Mawaheb I. Altuwaijri, Tarfa M. Albuhairi, Wejdan M. Alohaideb. Survey on Arabic Sentiment Analysis in Twitter.International Science Index, 9 (1), pp. 364-368, 2015.

[6] Korayem M, Crandall D and Abdul-mageed M. Subjectivity and sentiment analysis of arabic: a survey. In Advanced Machine Learning Technologies and Applications, , 128-139, 2012.

A. Farghaly and K. Shaalan. Arabic natural language processing: Challenges and solutions.ACM Transactions on Asian Language Information Processing (TALIP), volume 8(4): pages 1{22, Dec. 2009. ISSN 1530-0226.

[7] K. C. Ryding. A Reference Grammar of Modern Standard Arabic.Cambridge University Press, 2005.

[8] R.M. Duwairi, R. Marji, N. Sha'ban, S. Rushaidat, Sentiment analysis in Arabic tweets, Presented at the 5th international conference on Information and Communication Systems (icics), 2014, pp. 1–6.

[9] El-Halees, A.: Arabic Opinion Mining Using Combined Classification Ap-proach. In: Proceedings of the International Arab Conference on Information Technology, ACIT (2011).

[10] M. Rushdi-Saleh, M.T. Martin-Valdivia, L.A. Urena-Lopez, and J.M. Perea-Ortega, "Experiments with SVM to classify opinions in different domains". Expert Systems with Applications 38, 2011b, pp.14799- 14804.

A. Mountassir, H. Benbrahim, I. Berrada, an empirical study to address theproblem of unbalanced data sets in sentiment classification, Presented at theIEEE International Conference on Systems, Man, and Cybernetics (SMC), 2012, pp. 3298–3303.

B. Mourad, K. Darwish, Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs, Presented at the Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2013, pp. 55–64.

[11] M.A. Aly, A.F. Atiya, LABR: a large scale Arabic book reviews dataset, Presented at the ACL (2), 2013, pp. 494–498.

[12] S.R. El-Beltagy, A. Ali, Open issues in the sentiment analysis of Arabic social media: a case study, Presented at the 9th International Conference on Innovations in information technology (iit), 2013, pp. 215–220.

[13] M. Nabil, M. Aly, A.F. Atiya, ASTD: Arabic sentiment tweets dataset, Presented at the Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 2515–2519.

[14] ElSahar H, El-Beltagy SR. Building large Arabic multi-domain resources forsentiment analysis. In: Computational Linguistics and Intelligent TextProcessing.Springer; 2015. p. 23–34.

[15] H. Al-Rubaiee, R. Qiu, D. Li, Identifying Mubasher software products through sentiment analysis of Arabic tweets, Presented at the 2016 International Conference on Industrial Informatics and Computer Systems (CIICS), 2016, pp. 1–6.

[16] Al-Anzi, Fawaz S., and DiaAbuZeina. "Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing." Journal of King Saud University-Computer and Information Sciences 29.2 (2017): 189-195.

[17] Mariam Biltawi, WaelEtaiwi, Sara Tedmori, AmjadHudaib, and Arafat Awajan. 2016. Sentiment classification techniques for Arabic language: A survey. In 2016 7th International Conference on Information and Communication Systems (ICICS). IEEE, pages 339–346.

[18] Nawaf A Abdulla, Nizar A Ahmed, Mohammed A Shehab, and Mahmoud Al-Ayyoub. 2013. Arabic sentiment analysis: Lexicon-based and corpusbased. In 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). pages 1–6.

[19] Nawaf A Abdulla, Nizar A Ahmed, Mohammed A Shehab, Mahmoud Al-Ayyoub, Mohammed N Al-Kabi, and Saleh Al-rifai. 2016. Towards improving the lexicon-based approach for arabic sentiment analysis. In Big Data: Concepts, Methodologies, Tools, and Applications. IGI Global, pages 1970–1986.

[20] Song, Jihyun, Kyeongjoo Kim, and Minsoo Lee. "A Big Data Analysis and Mining Approach for IoT Big Data." International Journal 7.1 (2018).