ISSN 2007-9737

Using Tweets and Emojis to Build TEAD: an Arabic Dataset for Sentiment Analysis

Houssem Abdellaoui[1], Mounir Zrigui[2]

[1] National Higher Engineering School of Tunis: ENSIT, Research Laboratory LaTICE, Tunisia

[2] University of Monastir, Faculté des Sciences de Monastir, Tunisia

hsm.abdellaoui@gmail.com, mounir.zrigui@fsm.rnu.tn

Abstract. Our paper presents a distant supervision algorithm for automatically collecting and labeling 'TEAD', a dataset for Arabic Sentiment Analysis (SA), using emojis and sentiment lexicons. The data was gathered from Twitter during the period between the 1^{st} of June and the 30^{th} of November 2017. Although the idea of using emojis to collect and label training data for SA, is not novel, getting this approach to work for Arabic dialect was very challenging. We ended up with more than 6 million tweets labeled as Positive, Negative or Neutral. We present the algorithm used to deal with mixed-content tweets (Modern Standard Arabic MSA and Dialect Arabic DA). We also provide properties and statistics of the dataset alongside experiments results. Our tryouts covered a wide range of standard classifiers proved to be efficient for sentiment classification problem.

Keywords. Sentiment analysis, opinion mining, modern standard arabic, arabic dialect, sentiment dataset, emojis, sentiment lexicon.

1 Introduction

Sentiment analysis (SA), is the process of determining the sentiment or the opinion of a text. Obviously, we, as human beings, are good at this. We can look at a given text and immediately know what sentiment it holds (positive, negative or neutral). Companies and academic researchers across the world are trying to make machines able to do that. It is super useful for gaining insight into consumer's opinions.

Once you understand how your customers feel, after checking out their comments or reviews, you can identify what they like and what they don't, and build things for them such as, recommendation systems or more targeted marketing companies. The same logic can be applied in other fields for instance: economy, business intelligence, politics, sports, education and so on.

Prof. Lillian Lee (Cornell) is one of the founders of "Sentiment Analysis" as a field of study. It all started in the early 20^{th} century with his paper [20] and th work of Turney [29]. However, we can trace some few previous work related to SA such the research of Jaime [8] in 1979 that tackled the problem of subjectivity understanding and Ellen Spertus [28] who proposed a paper on automatic recognition of hostile messages in 1997.

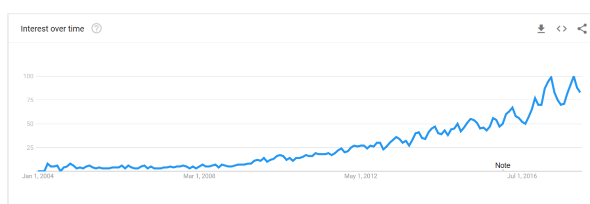

Fig. 1. Interest to SA from google trends (2004 - 2017)

Nowadays, SA is still gaining large attention. As shown in Figure 1, the trend of SA did not stop increasing since 2004. This is due to many facts. First, the evolution of Natural

Language Processing (NLP) is making a huge step towards understanding more and more the language computation and reasoning. Second, we have now much more computational power easily accessible than what we used to. Last but not least, the data is abundant on the web 2.0 especially on social networks like Twitter, Facebook, Instagram, etc. The work on SA is based on two main aspects. The first one focuses on creating algorithms and techniques (machine learning, lexicon based and linguistic). The second one is when researchers are trying to build linguistic resources such as datasets and lexicons for SA.

The work that we provide in this paper, follows the second aspect. So we are going to present the process that has been done to obtain a dataset for Arabic SA. We will discuss our approach to collect and label the dataset using emojis and sentiment lexicon. Also, we will highlight the problem of Arabic Dialect and how we managed to deal with it. Then, we will give details and statistics about the final TEAD dataset. And finally, we will conclude with the benchmark experiments and comparison with ASTD[18].

2 Related Work

The main goal of SA is detecting the polarity of a review. But it should be preceded by identifying the subjectivity to make sure that the expressed view is opinionated. For the polarity classification task, many datasets were suggested in the literature.

OCA [24] is one of the first sentiment datasets for Arabic language. It was manually collected from Arabic movies reviews. It contains 500 instances divided into 250 positives and 250 negatives. It served as a benchmark for many studies. In the same aspect LABR was proposed by Aly et.al[5] as the largest corpus for SA at that time. It holds more than 60 K review on books. The authors used a scale from one to five to rate them. Scale 1 and 2 for positive, 3 for neutral, 4 and 5 for negative. In 2012 Abdulmageed et al.[2] came up with AWATIF , a multi-genre corpus gathered from Wikipedia talk pages, web forms and Penn Arabic Tree Bank. AWATIF is not released online for free reuse or test.

ElSahar and El-Beltagy [11] collected a multidomain Arabic review dataset. The scope of the reviews included hotels, movies, products and restaurants. The role of social media is a key factor in the world where each part (corporations, brands, political figures, etc.) tries to have the most influence on users. The reasons behind this wave are simple. The first one, social media provides a huge amount of data easily accessible from users from all arround the world. Second, these contents are always there ready to be used freely. We just need to know how to mine it. Twitter is a micro-blogging website where users can share and send short text messages called tweets, limited to 140[1] characters.

It is thriving on the throne of social media with more than 6000 tweets per second. For the Arabic language; many pieces of research provide SA dataset collected from Twitter. Rafaee et al.[22] proposed a corpus for subjectivity analysis and SA. It comprises 6894 tweets (833 positives, 1848 negatives, 3685 neutrals and 528 mixed).Nabil et al. [18] used an automatic approach to construct their sentiment dataset. They called it ASTD; it consists of 10006 Arabic tweets divided into four classes (positive 793, negative 1684, mixed 832 and neutral 6691). Al-samadi [4] filtered LABR and selected 113 review. The selected ones were labeled for aspect-based SA. The annotation was made according to the SemEval2014-task4 guidelines.

In 2017, Nora [3] proposed AraSenti-Tweets [3] dataset of Saudi dialect with 17573 tweets manually labeled to four classes (positive negative neutral and mixed). Also, in the same year, the International Workshop on Semantic Evaluation proposed a new corpus for SA [23]. The data was gathered automatically from Twitter and manually labeled. The dataset was provided to SemEval participants to accomplish 5 tasks:

- *Subtask A.*: Message Polarity Classification: Given a message, classify whether the message is of positive, negative, or neutral sentiment.

- *Subtasks B-C.*: Topic-Based Message Polarity Classification: Given a message and a topic,

[1] The length of a tweet was expanded to 280 character starting from 26 Novembre, 2017

classify the message on B) two-point scale: positive or negative sentiment towards that topic C) five-point scale: sentiment conveyed by that tweet towards the topic on a five-point scale.

— *Subtasks D-E.*: Tweet quantification: Given a set of tweets about a given topic, estimate the distribution of the tweets across D) two-point scale: the ”Positive” and ”Negative” classes E) five-point scale: the five classes of a five-point scale.

Table 1. Arabic sentiment datasets

Authors	Dataset	Size	Source
[5]	LABR	63257	Goodreads
[2]	AWATIF	2855	Wikipedia, Forums
[4]	HAAD	1513	LABR
[24]	OCA	500	Arabic movies reviews
[3]	AraSenti-Tweets	17573	Twitter
[18]	ASTD	10006	Twitter
[23]	SemEval	8366	Twitter
[22]	–	6894	Twitter

3 The Need of Data and the Use of Emoji

3.1 Why Do We Need More Data?

How many instances do we need to train a sentiment classifier? The answer is not quite simple! No one can tell! This is an intractable problem that should be discovered through empirical investigation. The size of data required depends on many factors such as the complexity of the problem and the complexity of the learning algorithm. For example, if a linear algorithm achieves good performance with hundreds of examples per class, we may need thousands of examples per class for a nonlinear algorithm, like random the forest or deep neural networks.

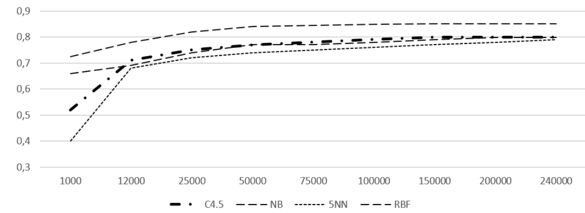

Fig. 2. The effect of dataset size on training tweet sentiment classifiers (RBF: radial basis function network,5NN: 5k-nearest neighbors, NB: naive Bayes, C4.5 algorithm) on English language [21]

Some studies tackled this problem like Pursa[21] and Kharde[15]. They concluded that dataset size significantly impacts classification performance as shown in figure 2. These two studies were carried out on the English language. Arabic is more complex compared to Latin languages due to its agglutinate nature. Each word consists of a combination of prefixes, stems, and suffixes that results in very complex morphology [1]. In fact, the SA task for Arabic needs much more data. Meanwhile, the literature review shows that freely available SA dataset for it are quite limited in size and number. In the remainder of this paper, we will present our tryouts to fill this gap by presenting TEAD an Arabic Sentiment Dataset collected from Twitter.

3.2 From Emoticons to Emoji to Sentiment Analysis

An emoticon is a stenography from facial expression. It eases the expression of feeling, mood, and emotion. It enhances written messages with some nonverbal elements that attract the attention of the reader and improves the overall understanding of the message. On the $19^{th}$of September 1982, Prof. Scott Fahlman of Carnegie Mellon University proposed the first emoticons. He used ”:-)” to distinguish jokes posts and ”:-(” for serious ones. After that, the use of emoticons had spread and new ones were created to express hugs, winks, kisses, etc.[14].

An emoji (Picture character in Japanese) is a step further. It appeared in Japan on the late 20^{th} century. It is used on modern communications technologies. It facilitates the expression of emotions, sentiments, moods and even activities. As a new ideogram, it represents more than facial expressions, but also ideas, concepts, activities, building cars, animals, etc.

Several studies analyzed the use and effect of emojis on social networks like Twitter. They showed that tweets, with emojis included, are more likely to express emotions [16][26][9]. Some other researches created an emojis lexicon for SA [19].

4 Data Collection and Pre-Processing

4.1 Collecting Data from Twitter

The process of gathering data for the training task was performed during the period between the 1^{st} of June and the 30^{th} of November 2017. Using Twitter API and an online server from OVH[2], we were able to collect thousands of tweets each day. We followed these steps:

— Select the top 20 most used emojis on Twitter according to emjoitracker[3] on the 31^{st} of May 2017.

— Use Sentiment Emoji Ranking [19] to choose the ones that are the most subjective (we ended up with 10 emojis presented in Table 2).

— Use Twitter Stream API V1.1[4] for tweets live streaming with 3 filters:

  - Language = Ar (for Arabic letters),
  - Contains = Filtered list of emojis,
  - Retweeted = No (to eliminate retweeted tweets).

The process yields to a dataset of 6 million Arabic tweets with a vocabulary of 602721 distinct entities.

[2] `https://www.ovh.com/`

[3] `http://emojitracker.com/`

[4] `https://blog.twitter.com/developer/en_us/a/2012/current-status-api-v1-1.html`

4.2 Arabic Scripts in Non-Arabic Languages

The Arabic script is not only used for writing Arabic, but also used in several other languages in Asia and Africa, such as Persian, Urdu, Azerbaijani, and others. Unfortunately, the Twitter stream-API is not able to detect whether the language of a tweet is Arabic or not. It was interesting to find how many tweets in other languages are there in our dataset. Sadly, we could not automate this process. We randomly extracted 2000 tweets and manually filtered them to find just one non-Arabic tweet. By this rate, we can assume that the exitance of such type of noise in our dataset TEAD is rare.

4.3 Translation From Arabic Dialect to MSA

Modern Standard Arabic (MSA), which is the official language of the Arab world, is not used as frequently as Arabic dialects in Web. Indeed, it is more used in newspaper articles, TV news, education or on official occasions, such as conferences and seminars. On a social network like Twitter, the use of dialect is very common[27].

However, since all the Arabic dialects have been using the same character set, and plenty of the terminology are shared among diverse varieties, it is not a minor matter to differentiate and discrete the dialects from each other. Although some studies [25] proposed machine learning methods to do that. We use the Twitter API to locate the origin of the tweet using geographic localization system. We divide the dataset into six groups. This partition is proposed by Sadat[25]. The results are in figure 3.

We used a simple and intuitive algorithm yet effective to replace dialect words with their respective synonyms in the MSA. The used dialect lexicons are presented in Table 3. Unfortunately, we were not able to find any lexicon for Iraqui dialect so we were forced to omit all the Iraqui tweets from our dataset. We also deleted the tweets from the class 'Others'.

ISSN 2007-9737

Table 2. List of the 10 most used Emojis on Twitter

Unified id	Emoji	Sentiment	Description
1F602	😂	Positive	FACE WITH TEARS OF JOY
2764	❤️	Positive	HEAVY BLACK HEART
1F60D	😍	Positive	SMILING FACE WITH HEART-SHAPED EYES
267B	♻️	Positive	BLACK UNIVERSAL RECYCLING SYMBOL
2665	♥️	Positive	BLACK HEART SUIT
1F62D	😭	Negative	LOUDLY CRYING FACE
1F60A	😊	Positive	SMILING FACE WITH SMILING EYES
1F612	😒	Negative	UNAMUSED FACE
1F629	😩	Negative	WEARY FACE
1F614	😔	Negative	PENSIVE FACE

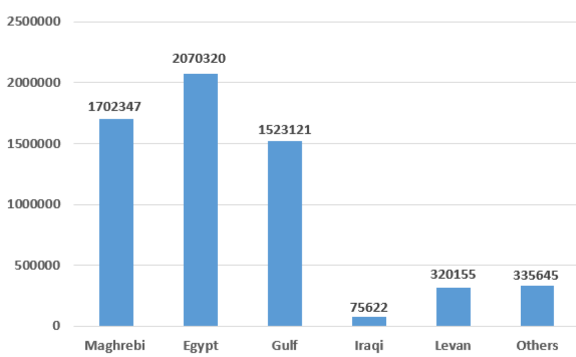

Fig. 3. Tweets repartition by dialect

Table 3. Lexicons used for translation of the Arabic dialect

Source	Dialect
[10]	Egypt
[12]	Levan
[13][27]	Maghrebi
[6]	Gulf

Table 4. TEAD dataset statistics

	Number of tweets	Average tokens per tweet	Max tokens per tweet
Positive tweets	3,122,615	9,42	45
Negative tweets	2,115,325	9,25	34
Neutral tweets	378,003	11,36	39

4.4 Manual Validation of the Automatic Annotation

4.5 Preprocessing

Preprocessing is an essential step in almost any NLP tasks. It aims to eliminate the incomplete, noisy, and inconsistent data. We followed this steps:

— **Removing URLs:** Tweets can contain links, so we need to remove them because they don't contribute to sentiment classification.

— **Removing usernames:** Usernames (@user) are also removed from the tweets.

ISSN 2007-9737

Table 5. Manual validation of the automatic annotation

	Number of tweets	Classification error rate
Positive tweets	1000	5,6%
Negative tweets	1000	4.2%
Neutral tweets	1000	11,3%

— **Remove duplicated letters:** We replaced any letter that appears consecutively more than two times in a word by one letter. For example the word جميييييلٌ *ğmyyyyyylun* becames جميلٌ *ğmylun*(beautiful) .

— **Remove punctuation and non Arabic symbols:** we also removed punctuation and others symbols that can be found in some tweets.

— **Tokenization and normalization:** we used stanford segmenter to perform the tokenization and normalization of the tweets.

4.6 Lexicon Based Approach for the Dataset Annotation

Facing the magnitude of the collected data, human labeling becomes expensive and takes a lot of time. We had to automate the process so we can keep up with the stream of tweets coming each minute. Our method to label the data is grounded on a lexicon-based approach for SA. It's detailed in Algorithm 1. It is worth mentioning that this algorithm can handle negation.

We used Ar-SeLn [7] the first publicly available large-scale Standard Arabic sentiment lexicon. We added the list of emojis used for gathering the tweets to the lexicons with their respective polarity according to the Sentiment Emoji Polarity Lexicon of Novak [19]. The result of the automatic annotation process is in Table 4.

To validate our automatic approach of the data annotation, we randomly extracted 1000 tweet from each class. We performed a manual labeling on these portions of data by 2 native Arabic speaking annotators. The classification error rate was satisfactory as shown in table 5. The highest value was on the neutral set (11%). This is due to the complexity of capturing the actual subjectivity of a tweet when the number of the positive tokens is equal to the negative ones.

5 Evaluation and Results

5.1 Technical Details

The training process aims to reveal hidden dependencies and patterns in the data that will be analyzed. Therefore, the training and test data set must be a representative sample of the target data. We conducted a set of benchmark experiments on TEAD and ASTD. Both datasets were randomly partitioned into training (70%) and test (30%). We used TF-IdF (token frequency inverse document frequency) and CBOW (continuous bag of words) as word representation features for classical ML algorithms.

For, the experiments using deep learning models, we used Word2vec [17] for word embedding. We trained Word2vec(Skip-gram) with optimal parameters (vector size= 300, min-count = 5, window =3). Experiments were coded in Python3.6 using Scikt-Learn[5] and Keras[6] with Google Tensorflow[7] as backend. We used a machine with AMD FX 6-Core (3.5 GHz) and 16 GB of RAM. Tensorflow used the NVIDIA CUDA Deep Neural Network library (cuDNN) v5.1 with Geforce GTX 940 as GPU.

5.2 Experimental Results and Discussion

From the experimental results, we can make the following observations:

[5]www.scikit-learn.org
[6]www.keras.io
[7]www.tensorflow.org

Algorithm 1: Sentiment Annotation Algorithm

```
Input  : List of positive tokens from Ars-SeLn Lp
Input  : List of negative tokens from Ars-SeLn Ln
Input  : List of all the tweets TEAD
Input  : List of all Arabic Negation words NegList
Output: List of positive tweets PosTweets
Output: List of negative tweets NegTweets
Output: List of objective tweets ObjTweets

begin
    foreach tweet tj of TEAD do
        SumPos/SumNeg accumulate the polarity of positive/negative tokens ;
        SumPos ← SumNeg ← 0;
        foreach word ti in tj do
            if ti in Lp and ti − 1 not in NegList then
                SumPos ← SumPos + 1;
            else if ti in Lp and ti − 1 in NegList then
                SumNeg ← SumNeg + 1
            else
                /* The word has no positive sentiment score in Ars-SeLn */
            end
            if ti in Ln and ti − 1 not in NegList then
                SumNeg ← SumNeg + 1;
            else if ti in Ln and ti − 1 in NegList then
                SumPos ← SumPos + 1
            else
                /* The word has no negative sentiment score in Ars-SeLn */
            end
        end
        if SumPos > SumNeg then
            PosTweets ⟵ PosTweets +{tj};
        end
        if SumNeg > SumPos then
            NegTweets ⟵ NegTweets +{tj};
        end
        if SumNeg == SumPos then
            ObjTweets ⟵ ObjTweets +{tj};
        end
    end
    return PosTweets, NegTweets,ObjTweets
end
```


doi: 10.13053/CyS-22-3-3031

ISSN 2007-9737

Table 6. Classification Experimental Results (in %) Using TF-Idf as Text feature extraction (SVM:Support vector machine , LR: Logistic regression, M-NB :Multinomial naive Bayes , B-NB : Bernoulli naive Bayes ,DT: Decision tree, RF: Random Forest).

	SVM		LR		M-NB		B-NB		DT		RF	
	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD
Precision	76	81	76	77	72	76	81	65	78	65	84	84
Recall	75	83	74	72	72	82	74	83	73	73	73	69
F1-score	75,5	**81,9**	74,9	74,4	74,4	76,6	74,9	81,9	68,7	75,4	68,7	75,7

Table 7. Classification Experimental Results (in %) Using CBOW as Text feature extraction (SVM:Support vector machine , LR: Logistic regression, M-NB :Multinomial naive Bayes , B-NB : Bernoulli naive Bayes ,DT: Decision tree, RF: Random Forest).

	SVM		LR		M-NB		B-NB		DT		RF	
	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD	ASTD	TEAD
Precision	70	88	71	75	73	70	86	66	73	66	85	85
Recall	79	82	80	81	72	83	79	82	75	70	75	47
F1-score	74,2	**84,8**	75,2	83,8	73,4	77,6	74,2	83,9	70,2	71,4	70,2	60,5

Table 8. Exerimental results precision using deep learning models.

Dataset	**LSTM**	**CNN**
ASTD	81%	79%
TEAD	**87.5%**	86%

— The hypothesis that we based our work on is: tweets with emojis are more likely to be subjective. As a matter of fact, the results of annotation algorithm in table 4 confirm the assumption. The tweets labeled as objective were much less than the subjective ones.

— The results of the classification task using traditional ML algorithms on our dataset TEAD outperformed the ones obtained using the ASTD dataset.

— We observe interesting patterns of correlation between training dataset size and learning results.

— SVM had the best experiment results and confirmed the previous work [18] assumption which is the suggested choice for SA.

— We used LSTM and CNN as deep learning (DL) classifiers. The less convenient results on ASTD proved that DL models need a huge amount of training data to achieve better results.

— LSTM trained on TEAD shows encouraging results and open the doors to further investigation for the use of such a model in Arabic SA task.

6 Conclusion and Future Work

In this paper we presented TEAD a large-scale Arabic tweets dataset. We provided details about the data collected. We used an emojis lexicon as keywords for data gathering and tried to overcame the problem of using dialect instead of MSA. Some of the benchmark experiments were established to compare TEAD to ASTD. Our dataset achieved a state of art performance with both classical ML and deep learning classifiers. It outperformed existing literature results. In future work we intend to:

— Increase the size of the dataset.

— Try to find a better approach to deal with Arabic dialect.

— Build a specific deep learning model for Arabic SA and train it on TEAD.

Acknowledgements

The authors are grateful to the referees and the editor for their constructive comments and helpful suggestions, without whom this work would not have been possible. The TEAD dataset is freely available online for academic and research purposes on {`https://github.com/HSMAabdellaoui/TEAD`}.

References

1. **AbdelRahman, S., Elarnaoty, M., Magdy, M., & Fahmy, A. (2010).** Integrated machine learning techniques for arabic named entity recognition. *IJCSI*, Vol. 7, pp. 27–36.

2. **Abdul-Mageed, M. & Diab, M. T. (2012).** Awatif: A multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. *LREC*, Citeseer, pp. 3907–3914.

3. **Abdulla, N. (2014).** *Towards building a sentiment analysis tool for colloquial and modern standard arabic reviews*. Ph.D. thesis, Master's thesis. Computer Science Department, Jordan University of Science and Technology, Irbid, Jordan.

4. **Al-Smadi, M., Qawasmeh, O., Talafha, B., & Quwaider, M. (2015).** Human annotated arabic dataset of book reviews for aspect based sentiment analysis. *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on*, IEEE, pp. 726–730.

5. **Aly, M. & Atiya, A. (2013).** Labr: A large scale arabic book reviews dataset. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pp. 494–498.

6. **Assiri, A., Emam, A., & Al-Dossari, H. (2018).** Towards enhancement of a lexicon-based approach for saudi dialect sentiment analysis. *Journal of Information Science*, Vol. 44, No. 2, pp. 184–202.

7. **Badaro, G., Baly, R., Hajj, H., Habash, N., & El-Hajj, W. (2014).** A large scale arabic sentiment lexicon for arabic opinion mining. *Proceedings of the EMNLP Workshop on Arabic Natural Language Processing (ANLP)*, pp. 165–173.

8. **Carbonell, J. G. (1979).** Subjective understanding: Computer models of belief systems. Technical report, YALE UNIV NEW HAVEN CONN DEPT OF COMPUTER SCIENCE.

9. **Dhaoui, C., Webster, C. M., & Tan, L. P. (2017).** Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*, Vol. 34, No. 6, pp. 480–488.

10. **Diab, M. T., Al-Badrashiny, M., Aminian, M., Attia, M., Elfardy, H., Habash, N., Hawwari, A., Salloum, W., Dasigi, P., & Eskander, R. (2014).** Tharwa: A large scale dialectal arabic-standard arabic-english lexicon. *LREC*, pp. 3782–3789.

11. **ElSahar, H. & El-Beltagy, S. R. (2015).** Building large arabic multi-domain resources for sentiment analysis. *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, pp. 23–34.

12. **Graff, D. & Maamouri, M. (2012).** Developing lmf-xml bilingual dictionaries for colloquial arabic dialects. *LREC*, Citeseer, pp. 269–274.

13. **Harrat, S., Meftouh, K., Abbas, M., & Smaili, K. (2014).** Building resources for algerian arabic dialects. *15th Annual Conference of the International Communication Association Interspeech*.

14. **Hogenboom, A., Bal, D., Frasincar, F., Bal, M., De Jong, F., & Kaymak, U. (2015).** Exploiting emoticons in polarity classification of text. *J. Web Eng.*, Vol. 14, No. 1&2, pp. 22–40.

15. **Kharde, V., Sonawane, P., et al. (2016).** Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.

16. **Kramer, A. D. (2012).** The spread of emotion via facebook. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 767–770.

ISSN 2007-9737

17. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

18. Nabil, M., Aly, M., & Atiya, A. (2015). Astd: Arabic sentiment tweets dataset. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2515–2519.

19. Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PloS one*, Vol. 10, No. 12, pp. e0144296.

20. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, pp. 79–86.

21. Prusa, J., Khoshgoftaar, T. M., & Seliya, N. (2015). The effect of dataset size on training tweet sentiment classifiers. *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*, IEEE, pp. 96–102.

22. Refaee, E. & Rieser, V. (2014). An arabic twitter corpus for subjectivity and sentiment analysis. *LREC*, pp. 2268–2273.

23. Rosenthal, S., Farra, N., & Nakov, P. (2017). Semeval-2017 task 4: Sentiment analysis in twitter. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 502–518.

24. Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A., & Perea-Ortega, J. M. (2011). Oca: Opinion corpus for arabic. *Journal of the American Society for Information Science and Technology*, Vol. 62, No. 10, pp. 2045–2054.

25. Sadat, F., Kazemi, F., & Farzindar, A. (2014). Automatic identification of arabic dialects in social media. *Proceedings of the first international workshop on Social media retrieval and analysis*, ACM, pp. 35–40.

26. Schuff, H., Barnes, J., Mohme, J., Padó, S., & Klinger, R. (2017). Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 13–23.

27. Sghaier, M. A. & Zrigui, M. (2017). Tunisian dialect-modern standard arabic bilingual lexicon. *Computer Systems and Applications (AICCSA), 2017 IEEE/ACS 14th International Conference on*, IEEE, pp. 973–979.

28. Spertus, E. (1997). Smokey: Automatic recognition of hostile messages. *AAAI/IAAI*, pp. 1058–1065.

29. Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics, pp. 417–424.

Article received on 12/01/2018; accepted on 05/03/2018.
Corresponding author is Houssem Abdellaoui.