ICIC Express Letters ICIC International ©2019 ISSN 1881-803X
Volume **13**, Number **10**, October **2019** pp. **899–906**

SENTIMENT ANALYSIS TO MONITOR STUDENT FEEDBACK IN THEIR NATIVE LANGUAGE

Mohammed Aldagdoog, Safaa Kother and Sadeq Alzaghair

Department of Information Technology
Directorate of Education
Hillah 51001, Babylon, Iraq
mo85th@gmail.com; mo85th@bab.epedu.gov.iq; { ssafaa1980; sadiq1985 }@yahoo.com

Received March 2019; accepted June 2019

Abstract. *In today's global education, students may be taught in a language, which is not native to them. If they are unable to articulate well in that language, then their feedback may not be clear. If we allow the students to express themselves in their native language, then we require that the lecturers also understand it. Sentiment analysis can address this problem by measuring the polarity (positive or negative) of opinion from their feedback. Equally, the system is a non-intrusive open text format, which allows for frequent measurement of the class mood. The system reported here considers a group of Iraqi students who provide feedback in the form of "Tweet"-like statements in Arabic. The results show some clear changes in group sentiment over time, while keeping a consistent spread.*
Keywords: Sentiment analysis, Arabic lexicon, Student feedback

1. **Introduction.** Obtaining useful feedback from students on the quality of a course is essential for improving that quality [1]. Many forms of feedback are used and are suitable for different situations [2,3]. Open-ended questions allow freedom of expression, but according to Richardson [1], "the burden of analyzing open-ended responses and other qualitative data is immense". With this in mind, the aim of this paper is to automate the analysis of open-ended feedback, using Sentiment Analysis (SA).

When students are from another country and the teaching language is not native to them, then requiring feedback in open-ended text may be difficult. By allowing them to respond in their own language, may also present problems for the lecturer. Translation tools could be one solution, but they do not always provide accuracy, as well as taking time to assess. An aim of this paper is to provide one effective way for foreign students to provide feedback in their own language. One secondary benefit is that feedback can be taken frequently in short "Tweet"-like statements, analyze and monitor the mood of the class throughout the course.

The sentiment of the student can obviously also be affected by other factors, beyond the lecturer and material. In this research, the time and the room were constant throughout. We cannot control individual's mood; however, by averaging over all the students, the impact of any significant changes in the sentiment of an individual should be reduced.

Sentiment analysis is the quantification of people's emotions and opinions from written text. For example, a business may want to know whether the public likes or dislikes its products from online reviews. With so many reviews, it is impossible to manually assess the overall sentiment. Instead, SA determines the emotions from the text and produces a single value, usually in the form of a positive or negative polarity.

DOI: 10.24507/icicel.13.10.899

Twitter posts [4,5] are relevant to this paper, since they are short (140 characters), similar to the feedback requested from the students. "Tweets" stimulate users to be concise which makes sentiments simpler and easier to assess. There are two main techniques for calculating sentiment; both of these have been used on Twitter posts. The first is a lexicon based approach, which determines the sentiment through the words and patterns used [6]. The second approach is supervised machine learning, which requires a set of known texts and their associated sentiments to be created as a training set. A model is then built using one of the many machine learning classification techniques (e.g., SVM (Support Vector Machine), Naïve Bayes) and used to classify the sentiment of a new text.

The lexicon approach is considered here. It requires only an Arabic Sentiment Lexicon – ArSenL [7]. The machine leaning approach though requires a training set. Currently there are no Arabic education training sets, although we will produce one at the end of this research, having processed a sufficient number of feedback texts.

The use of SA in the context of student feedback has been studied already by Munezero et al. [8] and Rodriguez et al. [9]. Both apply SA to academic texts (e.g., open-ended text responses to questionnaires and student learning diaries). Their definition of sentiment goes deeper to emotions such as joy, anger, sadness, surprise and fear. Their aim though is understanding individual's motivation and well-being. Our approach is much more focused, rapid and dynamic – capturing simple sentiments with hopefully greater accuracy than complex emotional states. Dataset was accumulated by collecting students' feedback using the FluidSurvey website. FluidSurvey questionnaire is a system designed to interact with people. It performs the following tasks: feedback forms, building questionnaires directly through the Internet, evaluations and calculating survey results.

We gathered data about two courses in one semester, by posting the questionnaire link to students via email. We have to ask two questions in the survey about students' satisfaction about the lecture and lecturer. Another way to obtain feedback is to manually distribute a paper questionnaire at the end of each lecture.

This paper contributes in building a new model (Arabic Sentiment Lexicon ArSenL) to assess the feelings of Arabic-speaking students about lectures they attend in English language due to the high accuracy of this approach. The reason behind choosing Arabic feedback is to allow the students to explain their sentiment in their mother tongue and this makes the feedback in a good quality. The other reason of choosing feedback in Arabic is that there is a few studies about Arabic language (more than 300 million speakers worldwide) due to the morphological richness that makes it challengeable in terms of applying sentiment analysis. Our system depends on a big and enhanced dictionary of Arabic words along with their scores (polarity).

2. **Related Work.** In recent years, sentiment analysis has become a significant interest for researchers attracting them to expand their studies in this vital field. In an early work, Liu [10] defined opinions as "subjective expressions that describe people's sentiment, appraisals or feelings towards entities, events and their properties". He differentiated between the comparative that elucidates the positivity or negativity of an item by matching it to other items and explicit opinions that give positive or negative statements about a certain theme.

Kim and Hovy [11] defined the opinion as a combination of four principles: sentiment, claim, topic and holder whereas Liu defined the explicit opinion as five principles: object, features, orientation, holder and time. In Kim and Hovy's perspective, a holder of an opinion claims a specific topic and relate his/her claim with a specific sentiment. On the other hand, in Liu's perspective, the holder discovers some features of specific objects and also relates a specific orientation about them at a specific time.

Lexicon approach was used by Qiu et al. [12] to identify the sentiment of sentences in contextual advertising. Their proposal allowed strategies of advertisement to increase relevance and improve user experience. Nasim et al. [13] classified English student feedback by employment of a hybrid approach that combines the use of machine learning methods and dictionary of sentiment in one unique system. Machine learning algorithms they used were SVM (Support Vector Machine) and Random Forest, while the sentiment lexicon was a modified version of MPQA (Multi Perspective Question Answering) that is a general-purpose lexicon.

Al-Ayyoub et al. [14] built a lexicon-based SA tool and a very large sentiment lexicon. They reported that the results they obtained showed that the proposed tool works very well on Arabic general tweets. Fernández-Gavilanes et al. [15] proposed an approach to anticipate sentiment in online textual messages such as tweets and reviews. Our system is not directly comparable to their system due to the difference in datasets and the novelty of their approach, they completely created sentiment lexicon PR40 (PolarityRank 40) using a total of 40 negative and positive seeds of words while we modified ArSenL.

Zhou et al. [16] proposed an expanded general lexicon with informal opinion expressions, opinion words, domain-dependent and abbreviations on Twitter. To apply it on general tweets, they improved SentiStrength (SS), which is a lexicon-based classifier that contains a slang list, a question word list, a general opinion word list, an idiom list, a negation word list, a booster word list and an emoticon list.

Our system is similar to a system developed by Abdulla et al. [17]. The most important similarity is size of the dataset; our student feedback is as the same as their tweets since it is concentrated set of few words. In addition, both studies recruited lexicon approach through building the word dictionary. Our research applied the approach on tweet like student feedback, while they analyzed Arabic tweets regarding many topics such as arts and politics. The current system modified ArSenL dictionary, which is lexicon-based approach to add more opinion words. The performance of a certain lexicon-based approach relies on the content of the lexicon dictionary, the big number of words in the dictionary and the noticeable good performance of the lexicon model. The method of words' polarity determination is the biggest challenge in the sentiment analysis. Some researchers used a combination of the lexicon-based and machine learning approaches all together in one system in order to enhance the performance of the classification process. In the system, Buckwalter translation was used to convert Arabic characters to ASCII due to the difficulty of the Arabic language. In addition ArSenL dictionary was built to calculate the score of sentiment.

3. **System Description.** RapidMiner platform was used to build the present system. It is an open source datamining tool and java-based environment and library. It also allows the user to access to the results and algorithms through a GUI. Moreover, RapidMiner supports Arabic text mining along with Python programming language as Python supports libraries of processing natural languages in particular strings.

Central to the system is the ArSenL database as shown in Figure 1, containing words and sentiments of formal, modern Arabic. Each feedback text is tokenized to a set of words and a search for a match takes place in the database. Each word in the database has two values representing a positive and negative polarity. Depending on the context, a word may have values for polarities. The values therefore represent its strength in both contexts. An additional processing step is included for negation. The system then sums these values across all the words in the text and proposes the polarity of highest value.

Tools such as NLTK and Scikit-learn were importantly used to implement the model we built especially preprocessing libraries Pandas and NumPy. After gathering student feedback, there must be a preprocessing before starting the sentiment analysis. Preprocessing has the following steps as shown in Figure 2.

Tokenization. The tokenization functions are included in NLTK libraries to divide documents into sentences, sentences into words, and words into letters in dependence on

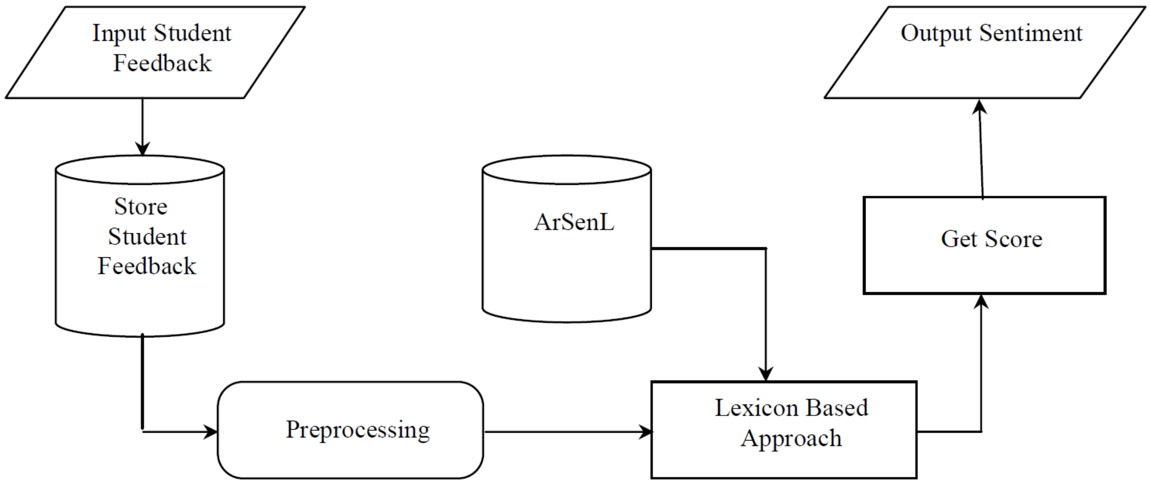

FIGURE 1. System architecture

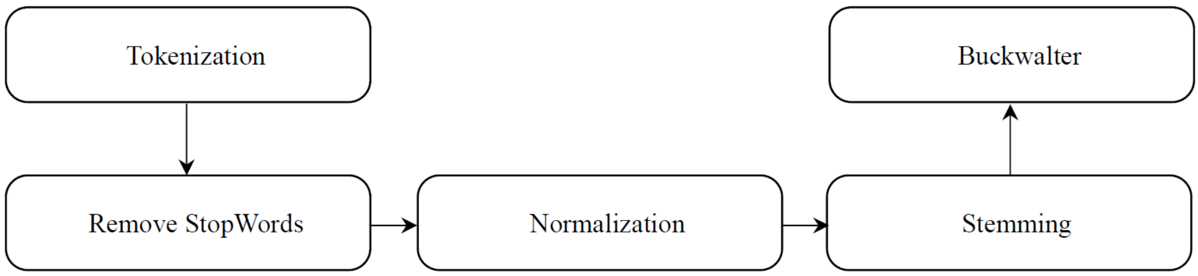

FIGURE 2. Preprocessing operations

the type of tokenization function. In this paper, word_tokenize function was used to divide the sentence into words because a student feedback is a short sentence.

Remove StopWords. StopWords is a set of words that repeatedly appear in the text such as (to, from, in, ...) and it is advisable to exclude them during the preprocessing of texts. In order to improve the accuracy, it is also recommended to ignore and not index them because removing stopwords does not impact the sentiment score of the certain sentence. The functions stopword.remove() and get_stop_words ('arabic') from the NLTK library were used to delete the Arabic stopwords in our system.

Normalization. Normalization process replaces interchangeable letters (similar letters) with one another; it also deletes the symbols from the sentences. Unlike the English language, Arabic language contains many short vowels (diacritics); therefore, normalization is important to remove such symbols like non-letters, punctuation, diacritics and Tatweel (elongation).

Stemming. Stemming process removes all suffixes and/or prefixes from words returning them back to their roots. It decreases the required size of the words to store them in indexes. Stemming process uses mathematical method to collect together all the words that share some semantic relations and have the same origin. Algorithms of Arabic stemming can be classified based on the root or based on origin according to the required level of analysis [18]. In this paper, Arabic stemmer ISRI (Information Science Research Institute) was used because our dictionary (ArSenL) uses ISRI stemmer to obtain the word root as well as it facilitates the matching process between student feedback and the dictionary.

Buckwalter. Buckwalter transliteration is the ASCII encoding of Arabic language characters [19]. Buckwalter is applicable in many commercial and academic purposes of natural language processing. It strictly converts one-to-one Arabic characters to ASCII characters; therefore, users may encounter difficulties in comprehending Buckwalter characters. In this paper, a piece of code was written in Python to function as conversion of Arabic symbols to ASCII characters according to Buckwalter transliteration method because ArSenL is coded using Buckwalter transliteration and Python neither handles Arabic characters like English characters nor supports them.

Arabic Sentiment Lexicon (ArSenL) was mainly built based on English SentiWordNet (ESWN), Standard Arabic Morphological Analyzer (SAMA), English WordNet (EWN) and most importantly Arabic Word-Net (AWN) because there is no direct lexicon for Arabic sentiment but there are many lexicons for English sentiment. The Arabic version (ArSenL) obtains its sentiment from the English one; however, ArSenL gets Arabic word sentiment from the English lexicon that corresponds to the Arabic words.

The main goal of this system is to find the final sentiment of the input sentences. When executing the operations, ArSenL receives word by word to perform the match process with list of words in the ArSenL. The sentiment score is positive or negative, if the match hits. The positive and the negative value are then stored. This process is iterative until it reaches the last word in the sentence. The word is ignored in case there is no match. After bringing all the negative and positive values of all words in a certain sentence, and storing them, the system sums up positive values of all words to obtain the final score of the sentence, and likewise of the negative sentences. Eventually, if the total score of the negatives is greater than the positives, then the sentiment is negative and vice versa as illustrated in the pseudo code for sentiment score calculation algorithm below. The system stores Arabic negation constructs such as (لا, كلا, ليس, ...). If one of these negation constructs appears in the sentence, then it flips the polarity of the following word.

Input: **S** student feedback (sentences in Arabic)
Output:
WP: word polarity, NP: number positive, NN: number negative
Initialization:
C = 0, positive, negative, pos, neg, Pos_sent, Neg_sent
Begin:
For each student feedback S, divide S into i word tokens Wi
    For each Wi do
    Remove stopword if found in Wi
    Arabic stemmer for each Wi
    Convert each letter in Wi to buckwalter mode
For each Entity in ArSenL
    Search Wi in ArSenL
If Wi $\in$ ArSenL then
Begin:
If Wi $\in$ negation then
    Wi+1
If Wi+1 $\in$ ArSenL then
    If pos in ArSenL > neg in ArSenL then
    pos = -1
    neg = 1
    else if pos in ArSenL < neg in ArSenL then
    pos = 1
    neg = -1
Else if Wi **not** in negation then
pos = pos of ArSenL, neg = neg of ArSenL
C++ as a counter to count how many Wi matching with ArSenL
End if

```
End if
If C equals 0 then
      pos = 0, neg = 0
Else
      pos = pos/C, neg = neg/C
Pos_sent+=pos, Neg_sent+=neg
End
If Pos_sent > Neg_sent
      output = positive
If Pos_sent < Neg_sent
      output = negative
Else
      output = neutral
End for
End for
End
```

The current system calculates the sentiment score of each student feedback. Two scores are there to represent the polarity of each sentence positive and negative. The student feedback is positive when the positive score is greater than the negative score and vice versa.

4. **Results.** Table 1 shows several text examples, their positive and negative scores as well as the resulting polarity. Example translations are made to demonstrate that the sentiment algorithm performs accurately on the Arabic texts. All polarity measures were checked manually found to be 100% consistent.

TABLE 1. Sentiment scores for Arabic feedback sentences

Negativity	Positivity	Polarity	Arabic Text
0.374	0.729	Positive	محاضرة اليوم كانت رائعة و سهلة جدا حيث كانت ذات فائدة كبيرة Toda's lecture was great, very easy and beneficial
0.139	0.355	Positive	تم شرح المحاضرة و دعمها بالأمثلة و فهمت بطريقة ممتازة The lecture was explained very well supported, with examples and understandable in an excellent manner
0.905	0.439	Negative	المحاضرة كانت جدا صعبة و لم افهم الموضوع بشكل كبير The lecture was very difficult and I did not understand the topic very well
0.357	0.685	Positive	المحاضرة كانت سهلة جدا و ممتعة، شكرا بروفسور The lecture was very easy and enjoyable, thanks professor
0.446	2.143	Positive	المحاضرة جيدة جدا من ناحية الفهم و الأستاذ جيد جدا و نستطيع القول ممتاز, نتمنى ان نحصل على معلومات مفيدة و نتمنى الوصول الى افضل مستوى The lecture was very good in terms of comprehension and the professor was very good too and we can say that he is excellent, we hope that we get useful information and reach the best level
0.842	0.21	Negative	المحاضرة صعبة جدا و لم افهم الموضوع The lecture was so difficult and I did not understand the topic

The trend and spread of sentiment over time can be seen for the two courses in Figure 3 and Figure 4 respectively. Each week reports the minimum and maximum scores and calculates the average. The three points (average, minimum and maximum) were processed using Microsoft Excel to plot each course over time.

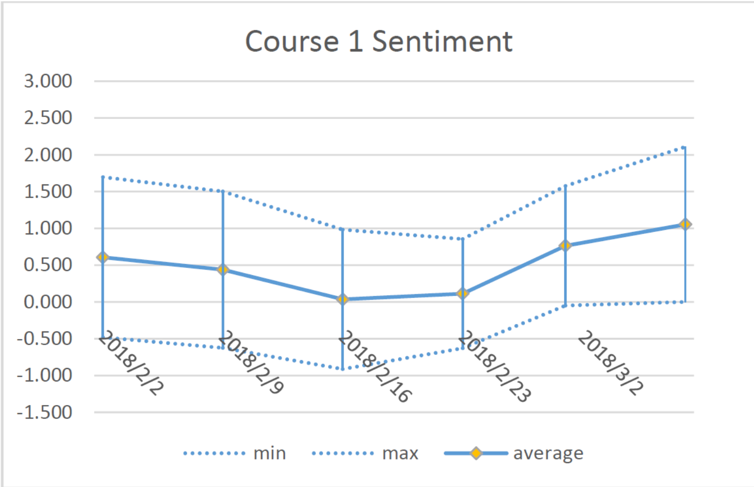

FIGURE 3. Sentiment average and spread over time (5 students)

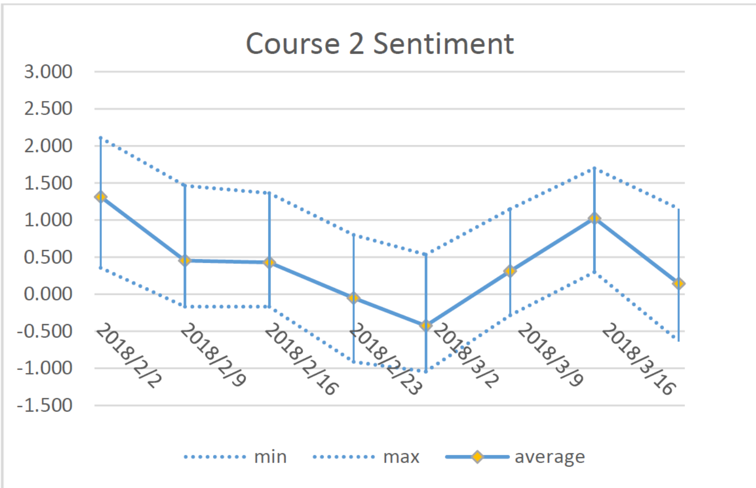

FIGURE 4. Sentiment average and spread over time (4 students)

There are several aspects to consider in these results. Course 1 appears more stable and remains positive throughout. Course 2 has a greater range of fluctuations and did descend to an average negativity in week 5, before recovering. Course 1 had a well-established curriculum with each lecture being self-contained. Course 2, however, was a new course, more technically difficult and had a greater dependency on what was learnt before. Its concepts took several weeks to become established with the students. Indeed, Course 2 included a lecture on sentiment analysis.

5. **Conclusion and Future Work.** The results seem plausible, considering the different styles of course. Both courses were considered successful, but whether these patterns persist, further experiments are needed. Certainly, for course 2 the intention is to avoid the descent to negativity by introducing easier, fun lectures within the first 4 weeks. This study has shown the potential of using social "Tweets" for additional student feedback.

By no means should the graphs always be positive, but successive negative sentiments or a large spread of sentiment, could trigger a review.

This system can be improved by including large and variant dataset such as general, product review, movie review, tweets, and political datasets. In addition, machine learning algorithms may be added to the current application in order to expand it and make comparisons between approaches.

REFERENCES

[1] J. T. E. Richardson, Instruments for obtaining student feedback: A review of the literature, *Assessment & Evaluation in Higher Education*, vol.30, no.4, pp.387-415, 2005.

[2] H. W. Marsh, Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research, *International Journal of Educational Research*, vol.11, no.3, pp.253-388, 1987.

[3] P. Ramsden and N. J. Entwistle, Effects of academic departments on students' approaches to studying, *British Journal of Educational Psychology*, vol.51, no.3, pp.368-383, 1981.

[4] H. Thakkar and D. Patel, Approaches for sentiment analysis on twitter: A state-of-art study, *arXiv preprint arXiv: 1512.01043*, 2015.

[5] R. M. Duwairi, R. Marji, N. Sha'ban and S. Rushaidat, Sentiment analysis in Arabic tweets, *The 5th International Conference on Information and Communication Systems*, pp.1-6, 2014.

[6] P. D. Turney and M. L. Littman, Measuring praise and criticism: Inference of semantic orientation from association, *ACM Trans. Information Systems (TOIS)*, vol.21, no.4, pp.315-346, 2003.

[7] G. Badaro et al., A large scale Arabic sentiment lexicon for Arabic opinion mining, *ANLP*, vol.165, 2014.

[8] M. Munezero et al., Exploiting sentiment analysis to track emotions in students' learning diaries, *Proc. of the 13th Koli Calling International Conference on Computing Education Research*, 2013.

[9] P. Rodriguez, A. Ortigosa and R. M. Carro, Extracting emotions from texts in e-learning environments, *The 6th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*, 2012.

[10] B. Liu, Sentiment analysis and subjectivity, in *Handbook of Natural Language Processing*, 2nd Edition, 2010.

[11] S.-M. Kim and E. Hovy, Determining the sentiment of opinions, *Proc. of the 20th International Conference on Computational Linguistics*, 2004.

[12] G. Qiu et al., DASA: Dissatisfaction-oriented advertising based on sentiment analysis, *Expert Systems with Applications*, vol.37, no.9, pp.6182-6191, 2010.

[13] Z. Nasim, Q. Rajput and S. Haider, Sentiment analysis of student feedback using machine learning and lexicon based approaches, *International Conference on Research and Innovation in Information Systems (ICRIIS)*, 2017.

[14] M. Al-Ayyoub, S. B. Essa and I. Alsmadi, Lexicon-based sentiment analysis of Arabic tweets, *International Journal of Social Network Mining*, vol.2, no.2, pp.101-114, 2015.

[15] M. Fernández-Gavilanes et al., Unsupervised method for sentiment analysis in online texts, *Expert Systems with Applications*, vol.58, pp.57-75, 2016.

[16] Z. Zhou, X. Zhang and M. Sanderson, Sentiment analysis on twitter through topic-based lexicon expansion, *Australasian Database Conference*, pp.98-109, 2014.

[17] N. A. Abdulla et al., Arabic sentiment analysis: Lexicon-based and corpus-based, *IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, 2013.

[18] T. M. T. Sembok and B. A. Ata, Arabic word stemming algorithms and retrieval effectiveness, *Proc. of the World Congress on Engineering*, vol.3, 2013.

[19] N. Habash, A. Soudi and T. Buckwalter, On Arabic transliteration, in *Arabic Computational Morphology*, Netherlands, Springer, 2007.