Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 05, No. 03, December 2019.

AUTOMATED ARABIC ESSAY GRADING SYSTEM BASED ON F-SCORE AND ARABIC WORDNET

Saeda A. Al Awaida[1], Bassam Al-Shargabi[1] and Thamer Al-Rousan[2]

(Received: 10-Jun.-2019, Revised: 29-Jul.-2019 and 20-Aug.-2019, Accepted: 24-Aug.-2019)

ABSTRACT

An Automated Essay Grading (AEG) system is designed to be used in universities, companies and schools, which depends on Artificial Intelligence and Natural Language Processing technologies; as it has the capability to improve the grading system in terms of overcoming cost, time and teacher effort while correcting the students' essay questions and papers. The AEG system widespread use is due to its cost, accountability, standards and technology; as that leads to the system being used and applied for multiple languages, such as English and French, among others. On the other hand, limited research has been conducted to automate Arabic essay grading. Therefore, this paper introduces an Arabic AEG system. In this paper, we propose a model for Arabic essay grading based on F-score to extract features from student answers and model answers along with the use of the Arabic WordNet (AWN) as a valuable knowledge-based method for semantic similarity. The purpose of using the AWN is to find all related words from student answers to give the answer a student score. Students will not be subject to injustice in terms of their marks in cases when they do not write the exact model answer, which subsequently leads to an improvement of the Arabic AEG system to match human grading. The proposed model is evaluated using Arabic essay dataset and the result shows that our proposed model produces a result which matches human grading.

KEYWORDS

Automated Essay Grading, Support Vector Machine, Arabic WordNet, Cosine Similarity, Natural Language Processing.

1. INTRODUCTION

Automated Essay Grading (AEG) techniques are used in grading student essays without the direct participation of individuals, where an AEG system can automatically evaluate and produce a score or grade for a written essay to tackle time, reliability and cost issues. AEG systems are motivated by the need to develop solutions to assist teachers in grading essays in an efficient and effective manner.

AEG systems keep on drawing the interest of government-funded schools, colleges, testing organizations, specialists and instructors. Numerous studies have been conducted to examine the accuracy and precision of AEG systems. Furthermore, there were several studies conducted on AEG systems which revealed high matching rates between AEG system scores and human scores with various AEG systems [1]. The idea of having compelling approaches and techniques to score essays of students is to liberate instructors of the burden of perusing and hand-scoring possibly hundreds of essays and papers.

Moreover, test publishers would most likely score essays and papers for a cheaper expense and possibly give higher-quality assigned scores by using the computer's special capabilities to improve AEG systems to achieve more accurate results compared to traditional scoring. AEG system mechanism contains many stages: collecting the student texts in text corpus form inputted into the AEG software. Firstly, the AEG system pre-processes the texts to make them useful for further processing and analysis. The basic pre-process technique includes stripping the texts of white spaces, removing certain characters such as punctuation, removing any character from other languages and splitting the text sequence into pieces, referred to as tokens. Other methods employed in the pre-processing will be illustrated in more detail in the next sections.

The second stage typically involves feature extraction, which is concerned with mapping the text sequence into a vector of measurable quantities, the most common examples and the frequency of each unique word in the text. It is considered as the most difficult part of the construction of an AEG system

1. S. A. Al Awaida and B. Al- Shargabi are with Middle East University, Amman, Jordan. E-mails: Saeda0awaida@gmail.com and bshargabi@meu.edu.jo
2. T. Al Rousan is with Isra University, Amman, Jordan. E-mail: thamer.rousan@iu.edu.jo

and it is challenging for humans to take into account all the factors affecting the grade. Furthermore, the effectiveness of the AEG system is constrained by the chosen features [2].

Many techniques were exploited to automate essay grading; techniques within the field of natural language processing, latent semantic analysis and machine learning [3]. Automated Arabic essay grading is still at its beginnings with limited research conducted in this field. In this direction, we propose in this paper a model to automate Arabic essay grading based on F-score to extract features from student answers and model answers along with the use of Arabic WordNet (AWN), which is a valuable system for semantic similarity measures and text similarity algorithms. The use of AWN is useful to find all related words from the student's answers, which match the meaning of such words in the model answer; in order to facilitate grading the students' essays. Students are not subject to injustice regarding their marks in cases when they do not write the same exact model answer, which subsequently leads to the improvement of the AEG system to match human grading.

2. LITERATURE REVIEW

This section presents an overview of the concepts and main topics of Arabic automated essay grading, which includes, F-score, AWN and the most recent related work.

2.1 F-score

Support vector machine is represented by sparse vectors under the vector space, where each word in the vocabulary is mapped into one coordinate axis. This is used on data to train a linear classifier which is characterized by the normal to the hyper-plane dividing positive and negative instances.

We apply feature selection aiming to pre-define the number of the highest scoring features to be included in a classifier by using the F-score technique. F-score is a feature selection technique in SVM. F-score measures the distinction between two classes (positive and negative), where each feature is assigned to a value computed as in [4]. If that value of F-score for the feature is bigger than the mean value of all F-scores in order for the feature to be added to feature space, the feature will not be considered for the feature space.

F-score is used in the proposed model to decide or select the feature, which affects the score of the student answer by determining the positive and negative classes according to a related or non-related answer, where the related answer (positive) takes it, while the others (negative) ignore it. F-score is good for feature selection, where it solicits each feature separately based on its score over the Fisher criterion, which prompts an optimal subset of features, especially when the features are extracted from text like essays to redundant features.

2.2 Text Similarity Algorithms

Many text similarity approaches have been used to develop automated essay grading systems [5]-[6]. There are three major approaches for text similarity: string-based similarities, corpus-based similarities and knowledge-based similarities, in addition to a sample of combinations of all of them. String-based similarities are also divided into two types; character-based and term-based, where these approaches measure similarity by counting the number of different characters in two sequences.

Corpus-based similarities are similarity measures between words based on information collected from a huge amount of texts which are mainly used for language research. The knowledge-based similarity is a semantic similarity measure, which relies on determining the ratio of similarity between texts using information collected from the semantic network. Moreover, some of these approaches are combined together to find optimal performance in terms of accuracy.

2.3 Arabic WordNet

Arabic WordNet is a valuable knowledge-based tool for several semantic similarity measures. It was created in 2006 and expanded in 2016 [7]. AWN is a lexical database for the Arabic language which is concerned with the meaning of words, rather than forms, where words are semantically similar. Moreover, its lexical resources contain not only words of the targeted language, but also synsets and semantic relations between words, such as synonymy, meronymy and antonymy; as synsets are groups

of words that can substitute other words in a sentence without changing its general meaning [7]. In this paper, we used Arabic WordNet to find all the related words from the student answers to give the student answer a score. Students are not subject to injustice regarding their marks when they do not write the same exact model answer.

2.4 Related Work

The latest and most recent related work is presented in this subsection, where many different approaches and systems were used to automate Arabic essay grading. An Arabic essay system called 'Abbir ' is presented in [8], where latent semantic analysis was used with some features, such as word stemming, spelling mistakes, proportion of spelling mistakes and word frequency, which revealed after different experiments that the performance is very close to human rating.

An automated assessor proposed in [7,9] for Arabic free text answer is based on LSA, which relies on replacing synonyms for each of the selected features to produce a matrix which is better than the traditional form of LSA matrix. The authors used the cosine similarity metrics to measure the similarity degree between the questions' model answers and the student answers. Accordingly, the score is given based on the higher ratio of similarity to set a score for the current essay based on model answer degrees.

A modified LSA is proposed in [9] for automatic essay scoring using Arabic essay answers, where a combined method of syntactic feature and LSA is based on bag-of-words. Afterwards, pre-processing creates a matrix, then applies cosine to define similarity. Results showed that syntactic feature improves accuracy. In this paper, we use AWN to apply the meaning features.

Moreover, a hybrid method employing LSA and rhetorical structure theory for automated Arabic essay scoring is proposed in [10], where the essay is semantically analyzed using LSA along with assessing essay writing style cohesiveness. The essay score of this approach was assigned based on the cohesion of the essay which represents 50% of the score, while 40% of the score is based on the writing style and the rest is given based on the spelling mistakes.

In [11], the authors suggested a web-based system which relies on using the Vector Space Model (VSM) to automate essay grading written in Arabic language. The system relies on two main processes; the first process extracts the features from essays, then applies SVM to find out the similarities between the essays written by the teachers and the ones written by the students, after converting each essay to vector space. The system then uses VS to match terms in the document after which cosine similarity is applied to find the score of student's answer. In this paper, SVM is used to extract features from answers. Another automated essay scoring system proposed in [12] also relies on using SVM, as it first extracts numerical features vector from the text data of essays using support vector machine classifier, then constructs a predictive model with extracted features and solves the multi-classification problem into multiple binary classifications to find the score between pairs of classes.

An approach using multiple classifiers, such as SVM, K-NN and Naïve Bayes in classifying Arabic language text documentation and comparing between those classifiers is used in [13]. The researchers used a dataset from Aljazeera news website and Al-Hayat website according to certain measures (recall, precision and F1), where the result suggests that the SVM classifier significantly outperforms other classifiers in high dimensional feature spaces. Accordingly, as the results in [13]-[14] indicate, F-score employed to extract features was the main reason of the model's high accuracy, which is why we used it in our proposed model.

A survey of similarity methods which focuses on challenges facing Arabic texts is employed in [15]. Three types of similarities were surveyed; lexical similarity based on character and statement similarities, semantic similarity and a hybrid similarity which combines both lexical and semantic similarities. The approach concluded that the cosine similarity metric produces an efficient performance when used in many Arabic essay grading systems, compared with other lexical measurements.

Moreover, a system based on the comparison of different text similarity algorithms for Arabic essay grading, such as string algorithm and corpus algorithm, is presented in [16]. The researchers applied multiple similarity measures to find an efficient way for essay grading. The N-gram approach is used in their system, as they relied on N-gram approach simplicity, which produces a reliable outcome when it comes to noisy data, such as grammatical errors or spelling mistakes, compared with the word-based approach.

A short answer system, based on translating student answers written in Arabic language into English language, is presented in [5] to tackle the challenges of the Arabic text in [6]. Accordingly, some problems existed during the translation process, such as a word in Arabic not in the same context structure and semantic is translated. Afterwards, the system applies multiple similarity measures and combines them to define the score of the tested student's answer. In this paper, we directly apply the similarity measure after extracting the feature without translation.

A system to automate essay scoring for online exams in Arabic language, based on evaluating the effects of stemming techniques, is applied in [17]. Heavy stemming and easy (light) stemming along with Levenshtein similarity measure are applied to the question in order to check the effectiveness of both techniques. As light stemming halts the elimination of prefixes and suffixes, without the ability to recognize the root of the word, heavy stemming is a root-based stemming which relies on eliminating prefixes and suffixes to get the actual root of a word. After finding the stemming word, the Levenshtein similarity measure is applied by giving each word a weight, then defining the distance between every two words to find the score.

3. PROPOSED MODEL

In this section, we present the proposed Arabic automated essay grading model, as illustrated in Figure 1, which is based on F-score, AWN and text similarity; to enhance the accuracy of the grading of essay exams. We developed a dataset (corpus) which is created to test the model. The proposed model consists of many phases, such as pre-processing, Arabic WordNet, feature extraction, using F-score and finally, applying the cosine similarity measure to determine the score of the student's answer; based on its cosine similarity degree and the model answers.

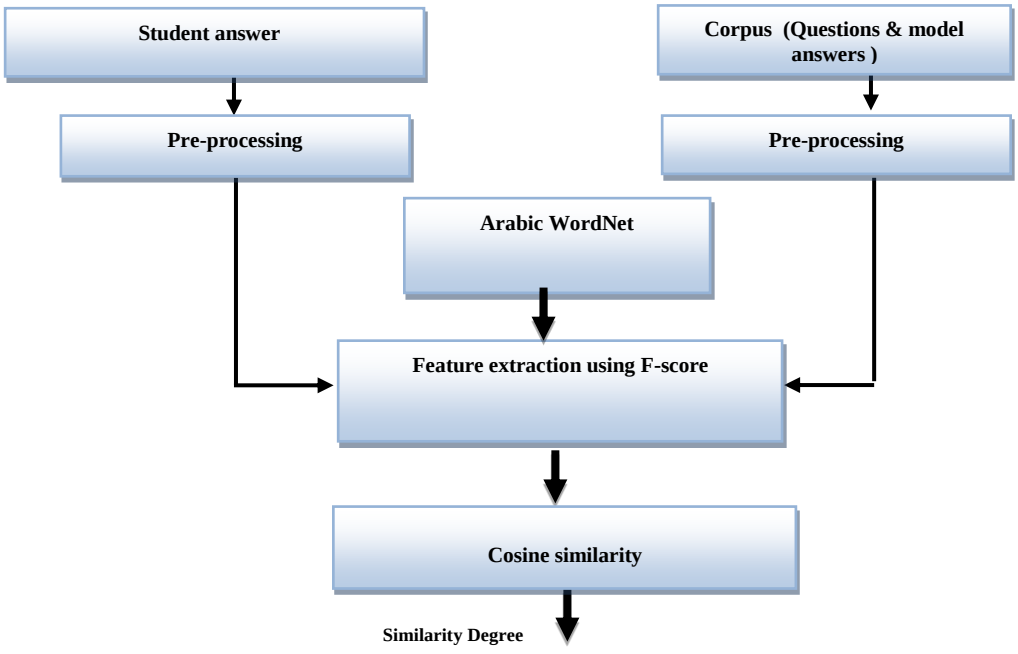

Figure 1. Proposed model.

3.1 Pre-processing

As seen in Figure 2, the pre-processing's first step is tokenization and normalization, which splits strings of student answer and model answer into smaller pieces and processes the transformation of the characters and the words into a single form. The second step is stop-word removal, where stop-words can be defined as words that do not have any significant meaning, or any word which does not have any importance and meaning in terms of finding the text classification; as these words are removed from the text. After converting the input Arabic text into a list of tokens, they are inputted to the next stage which is the stop-word removal, as they will be listed in the dictionary to remove them from the tokens output. The third stage is the stemming and lemmatizing process , which is a procedure's function for retrieving

the word to its basic root, by processing the removal of all of the word prefixes, suffixes and infixes. Lemmatization is closely related to stemming which extracts the base root of words. It creates an actual dictionary for words.

Example of the pre-processing step is as follows: An example of a student's answer:

مجموعة من الحواسيب والأدوات والمعدات ترتبط فيما بينها بوساطة خطوط اتصال

The result after the tokenization process:

""مجموعة" " من " " الحواسيب " "الأدوات" " ترتبط " " فيما " " بينها" " "المعدات" " بوساطة" " خطوط" " اتصال "

Stop word removal result:

"مجموعة" "الحواسيب " "المعدات" "الأدوات" " ترتبط " "بوساطة" "خطوط" "اتصال"

Different types of stemmers are used for Arabic text; in the proposed model, we used ISRI Arabic Stemmer to determine the roots of the Arabic words [18]-[19]. For the same example shown above, the ISRI stemming process is as follows :

"خط" "وصل" "جمع" "أداة" "عد" "حسب " " ربط " سط"

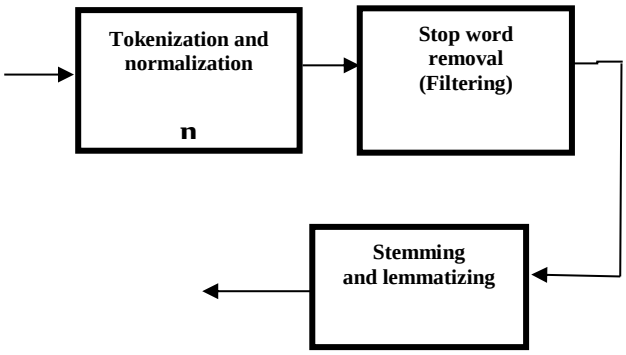

Figure 2. Pre-processing stages.

3.2 Arabic WordNet

As mentioned earlier, the WordNet is a lexical database which groups the words into sets of synonyms called synsets, along with the relations among these synonym sets, as finding a lexical resource offers broad coverage of the general lexicon of each word in the student's answer which is extracted from the previous stage to define all the words that have a similar meaning. We have used the AWN which is a multi-lingual concept dictionary that maps between word senses in Arabic and those in the Princeton WordNet that was expanded in 2016 [7]. We used the AWN in this research to find all the words that are synonymous with the student's answer; to increase the likelihood of the student's correct answer which was used after the pre-processing step. For the same example shown above , the result of using AWN is shown in Table 1.

Table 1. Example of WordNet.

word	synonyms
وصل	بلغ , علم , نهى .....
وسط	قصد, جزع, قطع .....
ربط	وثق , شد .....
جمع	حقن, قرن , لصق, ألف, حفظ, حشد,.....

3.3 F-score for Feature Selection

The support vector machine is represented by sparse vectors under the vector space, where each word in the vocabulary is mapped to one coordinate axis. It is used on data to train a linear classifier which is characterized by the normal to the hyperplane dividing positive and negative instances [20], [21], [22]. We apply feature selection aiming to pre-define the number of the highest scoring features to be included in a classifier by using the F-score technique. F-score is a feature selection method in SVM, which identifies the differences between two classes (positive and negative). The value of F-score for each feature is computed using Equation (1) [4]:

$$F(i) = \frac{\left(\overline{x_i^+} - \overline{x_i}\right)^2 + \left(\overline{x_i^-} - \overline{x_i}\right)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}\left(x_{k,i}^{(+)} - \overline{x_i^{(+)}}\right)^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}\left(x_{k,i}^{(-)} - \overline{x_i^{(-)}}\right)^2} \quad (1)$$

where *K* is a positive or negative instance, *xi* ¯, *xi* +: the average of i feature positive and negative dataset. *k*, *i*: the jth feature of the ith positive /negative instance. After determining the score of each feature, we then obtain the threshold value through calculating the average of F-score for all features. If the value of F-score is bigger than the mean value of all F-scores, the feature is added to the feature space, whereas if the value of F-score is less than the mean value of all F-scores, the feature is removed from the feature space. F-score is used to decide or select the feature that affects the score of student's answer which determines the positive and negative; according to a related or un-related answer. As shown in Figure 3, the related answers (positive) are taken, but the others (negative) are ignored.

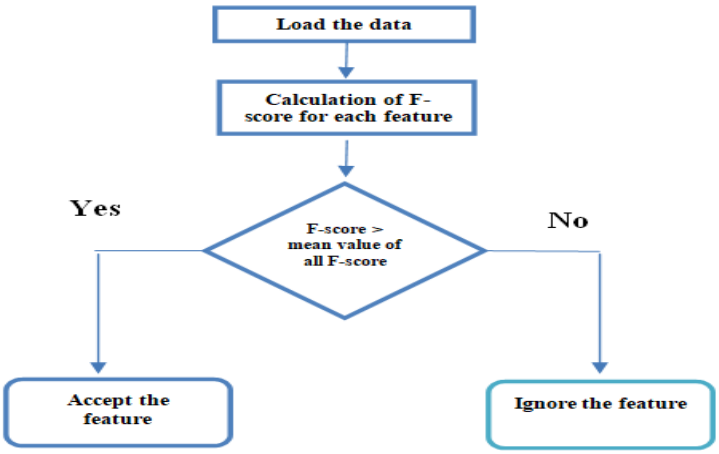

Figure 3. F-score feature selection.

3.4 Cosine Similarity

Cosine similarity is used to measure the cosine of the angle between any two vector spaces. It can be seen as "a comparison between documents" on a normalized space, as we are not taking into account the weight of each word count for each document, but the angle between documents. Cosine similarity will generate a value which articulates how correlated the two documents are by considering the angle as an alternative of the magnitude [7,22]. Cosine similarity is computed using Equation (2):

$$\cos\theta = \frac{\sum_i w_{q,i} . w_{i,j}}{\sqrt{\sum_j w_j^2} \cdot \sqrt{\sum_j w_{i,j}^2}} \quad (2)$$

4. Experimental Design and Results

To evaluate the proposed model effectively, we carried out in this paper a comparative analysis of the impact of Arabic WordNet in automated essay grading. The dataset used is created in MYSQL as a CSV file, as data is collected from a computer, science and social school lectures from Allu'lu'a modern

school in Madaba-Jordan, containing 120 questions (a sample of the questions used is shown in Table 2) along with 3 classes of model answers for each question and for each question with 30 student sample answers, a sample of model answers with human score is shown in Figure 4. The dataset was designed after the Hewlett Foundation Automated student assessment prize. The experiments were divided into two stages, where the first stage is the use of the proposed model without AWN; the cosine similarity degree for our proposed model and the human score are shown in Table 3.

Table 2. Sample of questions in the dataset.

Question ID	Question Text
1	عرف الإنترنت
2	ماذا يقصد بأمن المعلومات؟
3	ماذا يقصد بمزود خدمة الإنترنت؟
4	ماذا نعني بمحركات البحث؟
5	ما هي متطلبات الاتصال بالإنترنت؟
6	ماذا يقصد بمزود خدمة الإنترنت؟
7	ما هي وظيفة جهاز المودم؟
8	اذكر خدمات البريد الإلكتروني
9	علل احترام حقوق الملكية الفكرية عند استخدام الإنترنت؟
10	عرف بروتوكول الإنترنت
11	عرف الصراع الاجتماعي
12	عرف المنصهر
13	عرف البوصلة
14	عرف الكوكب
15	اذكر مكونات الدارة الكهربائية

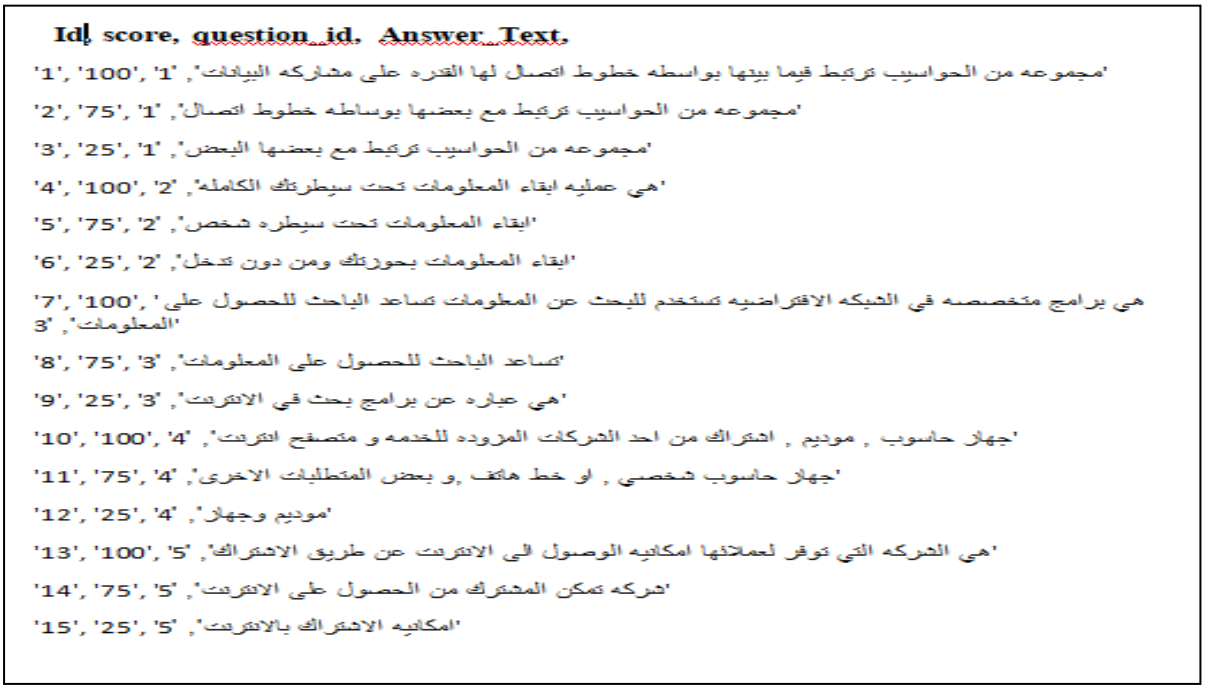
Id, score, question_id, Answer_Text,
'1', '100', '1', 'مجموعه من الحواسيب ترتبط فيما بينها بواسطه خطوط اتصال لها القدره على مشاركه البيانات',
'2', '75', '1', 'مجموعه من الحواسيب ترتبط مع بعضها بوساطه خطوط اتصال',
'3', '25', '1', 'مجموعه من الحواسيب ترتبط مع بعضها البعض',
'4', '100', '2', 'هي عمليه ابقاء المعلومات تحت سيطرتك الكامله',
'5', '75', '2', 'ابقاء المعلومات تحت سيطره شخص',
'6', '25', '2', 'ابقاء المعلومات بحوزتك ومن دون تدخل',
'7', '100', 'هي برامج متخصصه في الشبكه الافتراضيه تستخدم للبحث عن المعلومات تساعد الباحث للحصول على 3', 'المعلومات',
'8', '75', '3', 'تساعد الباحث للحصول على المعلومات',
'9', '25', '3', 'هي عباره عن برامج بحث في الانترنت',
'10', '100', '4', 'جهاز حاسوب , موديم , اشتراك من احد الشركات المزوده للخدمه و متصفح انترنت',
'11', '75', '4', 'جهاز حاسوب شخصي , او خط هاتف ,و بعض المتطلبات الاخرى',
'12', '25', '4', 'موديم وجهاز',
'13', '100', '5', 'هي الشركه التي توفر لعملائها امكانيه الوصول الى الانترنت عن طريق الاشتراك',
'14', '75', '5', 'شركه تمكن المشترك من الحصول على الانترنت',
'15', '25', '5', 'امكانيه الاشتراك بالانترنت',

Figure 4. Sample of model answers.

The second stage of the proposed model uses the AWN; the cosine similarity degree for our proposed model and the human score are shown in Table 4.

Table 3. Result of the proposed model without WordNet.

Question(id)	human score	cosine result without WordNet
1	1	0.94
2	1	0.94
4	0.75	0.67
5	0	0
11	0.75	0.66
12	0.75	0.82
15	0.25	0.21
30	0.25	0.22
36	0.25	0.1
40	1	0.97

Table 4. Score result of the proposed model using WordNet.

Question(id)	human score	cosine result with WordNet
1	1	0.98
2	1	0.98
4	0.75	0.8
5	0	0
11	0.75	0.67
12	0.75	0.85
15	0.25	0.24
30	0.25	0.3
36	0.25	0.21
40	1	0.98

The result of experiments demonstrates that the accuracy of the proposed model with AWN is close to human score as illustrated in Figure 5.

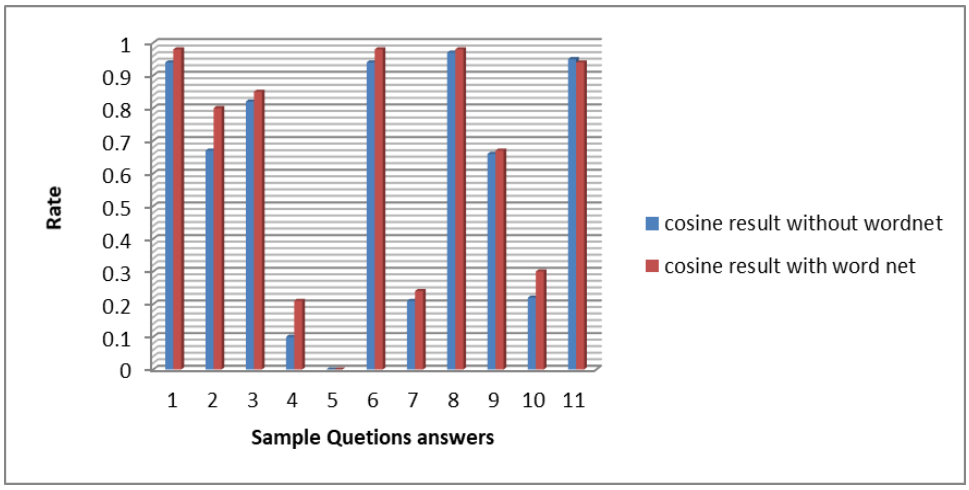

Figure 5. Rates of cosine similarity with and without AWN.

Moreover, to evaluate the effect of using AWN in the proposed model, a comparison was conducted between the human score and the score produced by the proposed model for a student's answer, using Mean Absolute Error (MAE) value and Pearson Correlation Coefficient. The MAE of the proposed model with the use of Arabic WordNet is 0.117 less than MAE of the proposed model without using Arabic WordNet. So, this result indicates that the proposed model will improve in the Arabic Automated Essay System with AWN, as shown in Figure 6.

Moreover, we used Pearson correlation which is a statistical measure that is used to determine whether or not there is a correlation between the scores produced by the proposed model and the human score.

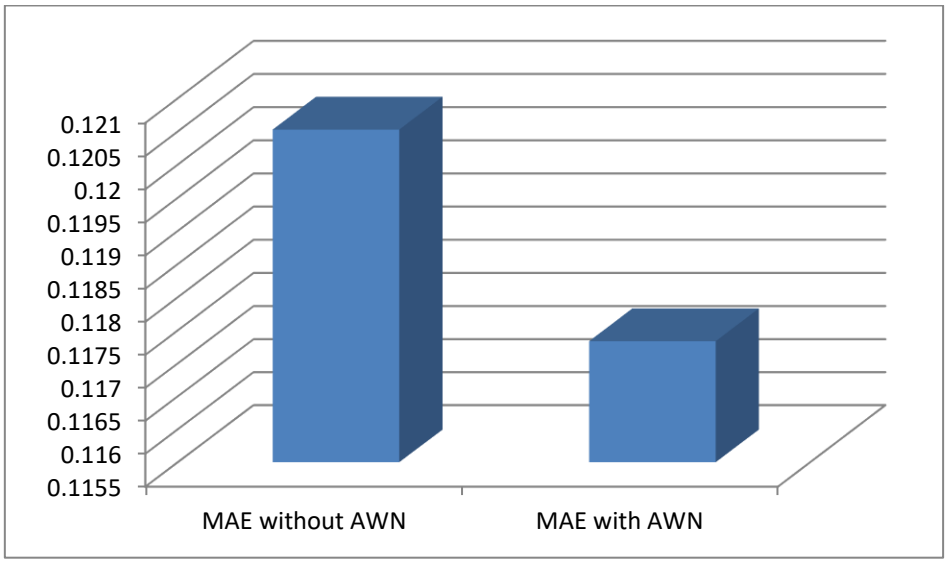

Figure 6. MAE.

Accordingly, the Pearson correlation result for the proposed model compared to human score is between 0.5 and 1, as it shows a high positive correlation that represents having the best correlation magnitude, as shown in Figure 7.

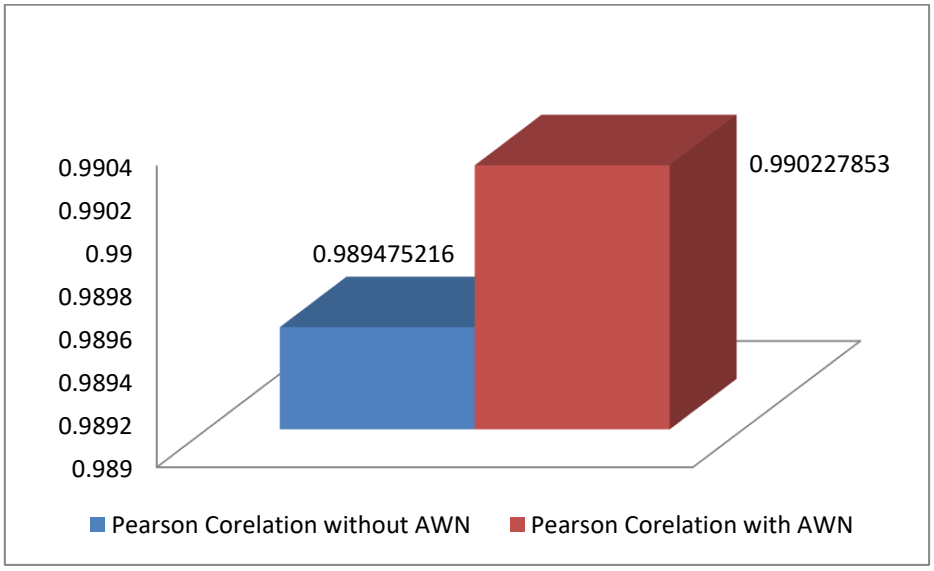

Figure 7. Pearson correlation.

5. Conclusion

This paper presented an Automated Arabic essay grading model to achieve better accuracy, while studying the role of Arabic WordeNet and F-score as efficient tools to extract features from the students' answers and from the model answers. We used cosine similarity to compute the score for the students. The focus of this work was to enhance the accuracy of the automated essay system to match human score by adding Arabic WordNet. The dataset created contains 120 questions with 3 model answers for each question. In addition, the dataset used in this paper was created according to Hewlett Foundation ASAP standards from Kaggle datasets. The automated essay grading results showed that the proposed model coupled with using Arabic WordNet (AWN) produces a better result compared to the case without using Arabic WordNet according to mean absolute error value and Pearson correlation. Based on the outcome of this research, there is an outlook for future work on using machine learning and neural network models to enhance the accuracy of Arabic essay grading along with studying the impact of the word-embedding technique.

REFERENCES

[1] D. Surya, V. Madala, A. Gangal, S. Krishna, G. Anjali and S. Ashish, "An Empirical Analysis of Machine Learning Models for Automated Essay Grading," PeerJ Preprints, pp. 1-14, India, 2018.

[2] V. Ramalingam, A. Pandian, P. Chetry and H. Nigam, "Automated Essay Grading Using Machine Learning Algorithm," Journal of Physics, Conference Series, vol. 1000, p. 012030, [Online], Available: 10.1088/1742-6596/1000/1/012030, 2018.

[3] M. Lilja, Automatic Essay Scoring of Swedish Essays Using Neural Networks, Statistics Uppsala University, 2018.

[4] A. Suresh and M. Jha, "Automated Essay Grading Using Natural Language Processing and Support Vector Machine," International Journal of Computing and Technology, vol. 5, no. 2, 2018.

[5] S. Gunes, K. Polat and S. Yosunkaya, "Multi-class F-Score Selection Approach to Classification of Obstructive Sleep Apnea Syndrome," Expert Syst. Appl., vol. 37, pp. 998-1004, 2010.

[6] W. H. Gomaa and A. A. Fahmy, "A Survey of Text Similarity Approaches," International Journal of Computer Applications, vol. 68, no. 13, 2013.

[7] Y. Regragui, L. Abouenour, F. Krieche, K. Bouzoubaa and P. Rosso, "Arabic WordNet: New Content and New Applications," Proceedings of the 8th Global WordNet Conference, pp. 330-338, 2016.

[8] A. Ewees, A. Mohammed Eisa and M. M. Refaat, "Comparison of Cosine Similarity and kNN Automated Essay Scoring," International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 13, pp. 8669-8673, 2014.

[9] M. Alghamdi, M. Alkanhal, M. Al-Badrashiny, A. Al-Qabbany, A. Areshey and A. Alharbi, "A Hybrid Automatic Scoring System for Arabic Essays," AI Communications, vol. 27, no. 2, pp. 103-111, 2014.

[10] R. Mezher and N. Omar, "A Hybrid Method of Syntactic Feature and Latent Semantic Analysis for Automatic Arabic Essay Scoring," Journal of Applied Sciences, vol. 16, no. 5, 2016.

[11] M. F. Al-Jouie and A. M. Azmi, "Automated Evaluation of School Children Essays in Arabic," Proc. of the 3rd International Conference on Arabic Computational Linguistics, vol. 117, pp.19-22, 2017.

[12] A. R. Abbas and A. S. Al-qaza, "Automated Arabic Essay Scoring (AAES) Using Vector Space Model (VSM)," Journal of Al-Turath University College, vol. 15, pp. 25-39, 2014.

[13] R. Martinez, H. H. Dong and D. Lee, "Automated Essay Scoring System by Using Support Vector Machine," International Journal of Advancements in Computing Technology (IJACT), vol. 5, no. 11, pp. 316-322, 2013.

[14] T. F. Gharib, M. B. Habib and Z. T. Fayed, "Arabic Text Classification Using Support Vector Machines," International Journal of Computers and Their Applications, vol. 16, no. 4, pp. 192-199, 2009.

[15] S. Alsaleem, "An Automated Arabic Text Categorization Using SVM and NB," International Arab Journal of e-Technology, vol. 2, no. 2, pp. 124-128, 2011.

[16] S. S.Aljameel, J. D. O'Shea, K. A. Crockett and A. Latham, "Survey of String Similarity Approaches and the Challenging Faced by the Arabic Language," Proceedings of 11th International Conference on Computer Engineering and Systems, Cairo, Egypt, 2016.

[17] A. Shehab, M. Faroun and M. Rashad, "An Automatic Arabic Essay Grading System Based on Text Similarity Algorithms," International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018.

[18] E. F. Al-Shalabi, "An Automated System for Essay Scoring of Online Exams in Arabic Based on Stemming Techniques and Levenshtein Edit Operations," arXiv preprint, 2016.

[19] B. Al-Shargabi, W. Al-Romimah and F. Olayah, "A Comparative Study for Arabic Text Classification Algorithms Based on Stop Words Elimination," Proceedings of the International Conference on

Intelligent Semantic Web-services and Applications (ACM), 2011.

[20] B. Al-Shargabi, F. Olayah and W. W. Romimah, "An Experimental Study for the Effect of Stop Words Elimination for Arabic Text Classification Algorithms," International Journal of Information Technology and Web Engineering (IJITWE), vol. 6, no. 2, pp. 68-75, 2011.

[21] G. Bilal and N. Rasha, "Semantic Analysis Based Customer Reviews Feature Extraction," Journal of Babylon University/Pure and Applied Sciences, vol. 25, no. 3, 2017.

[22] M. H. Nnguyen and F. Torre, "Optimal Feature Selection for Support Vector Machine," Pattern Recognition, vol. 43, pp. 584–591, 2014.

ملخص البحث:

في هذا البحث، يتم تصميم نظام مؤتمت لوضع الدرجات لاستخدامه في الجامعات والشركات والمدارس، يعتمد على الذكاء الاصطناعي وتقنيات المعالجة القائمة على اللغات الطبيعية. ويتميز النظام بقدرته على تحسين نظام وضع الدرجات من حيث التغلب على المشكلات المتعلقة بالتكلفة والوقت والجهد المبذول من المعلّم عندما يقوم بتصحيح إجابات الطلبة وأوراقهم. ويعود السبب في الانتشار الواسع لأنظمة وضع الدرجات المؤتمتة الى أمور تتعلق بالتكلفة والموثوقية والمعايير والتقنيات، وهي أمور تجعل النظام صالحاً للتطبيق والاستخدام لنصوص بلغاتٍ متعددة مثل الإنجليزية والفرنسية، ولغات أخرى. من جهة أخرى، لم يتم سوى إجراء عدد محدود من البحوث والدراسات في مجال أتمتة وضع الدرجات للنصوص العربية.

لذا، تقترح هذه الورقة نظام مؤتمت لوضع الدرجات لنصوص اللغة العربية. ويرتكز النظام المقترح على استخدام درجة ف (F-score) لاستخراج الخصائص من إجابات الطلبة ومن الإجابات النموذجية، جنباً الى جنب مع استخدام شبكة "ووردنت" العربية (AWN) كطريقة قيّمة مبنية على المعرفة من أجل التشابه في دلالات الألفاظ. ويتمثل الغرض من استخدام الشبكة العربية في إيجاد جميع الكلمات ذات العلاقة من إجابات الطلبة لإعطاء الطالب درجة على إجابته. وفي هذه الحالة، لن يتعرض الطلبة للإجحاف بدرجاتهم في الحالات التي لا يكتبون فيها الإجابة النموذجية الدقيقة؛ الأمر الذي من شأنه أن يقود الى تحسين نظام وضع الدرجات بحيث يتلاءم مع وضع الدرجات اليدوي الذي يقوم به الإنسان. وقد جرى تقييم النموذج المقترح في هذه الدراسة باستخدام مجموعة بيانات لنصوص باللغة العربية. وبينت النتائج أن النموذج المقترح يعطي نتائج تتوافق مع وضع الدرجات الذي يقوم به الإنسان.