Textual Entailment for Arabic Language based on Lexical and Semantic Matching
Mariam Khader, Arafat Awajan and Akram Alkouz
Pages 67 – 74
DOI: http://dx.doi.org/10.21700/ijcis.2016.109

Textual Entailment for Arabic Language based on Lexical and Semantic Matching

Mariam Khader, Arafat Awajan, Akram Alkouz
Department of Computer Science, Princess Sumaya University for Technology
[mkhader, awajan, akram]@ psut.edu.jo

Abstract: *Textual Entailment is one of the recent natural language processing challenges, where the meaning of an expression "Hypothesis" could be entailed by the meaning of another expression "Text". In comparison with English Language, the Arabic language has more challenges in determining entailment judgment, due to lexical and syntactic ambiguity. The proposed work in this paper has adopted a lexical analysis technique of Textual Entailment to study the suitability of this technique for Arabic language. In addition a semantic matching approach was added in order to enhance the precision of the proposed entailment system. The lexical analysis is based on calculating word overlap and bigram extraction and matching. The semantic matching has been incorporated with word overlap to increase the accuracy of words matching. The system has been evaluated by measuring precision and recall, those two metrics are the main evaluation measures used in Recognizing Textual Entailmen challenge number 2 to evaluate participated systems. The system has achieved precision of 68%, 58% for both Entails and NotENtails respectively with overall recall of 61%.*

Keywords: *Textual Entailment; Lexical Analysis; RTE Challenge; Semantic Matching; Arabic Natural Language Processing.*

Received: May 01, 2016 | ***Revised:*** August 10, 2016 | ***Accepted:*** September 14, 2016

1. Introduction

Textual entailment is one of the new challenges in Natural Language Processing (NLP), that has been proposed in 2004 by Dagan and Glickman [1]. Textual Entailment is defined as the process of determining, given two text expressions, if the meaning of one expression of the text (Text) could be inferred from the second text expression (Hypothesis) [2] [3]. This inference could be employed in various applications that need to find if a sentence infer another, such as: Information Extraction (IE), Machine Translation Evaluation, Language Generation Tools, Automatic Summarization, Answer Validation, Text Mining, Paraphrases (PP) and Information Retrieval (IR) [2] [3].

The Recognizing Textual Entailment (RTE) challenge, which has been announced in 2005, has encouraged researchers to develop and improve Textual Entailment systems. The Textual Entailment became the research problem for many studies and main discussion in several workshops and scientific meetings, such as: Semantic Text Processing Symposium, Joint Symposium on Semantic Processing (Textual Inference and Structures in Corpora) and many others [4]. RTE challenges have emphasized the need for further research in this field towards improving the accuracy of developed systems and applying Textual Entailment for other languages than English, such as Arabic.

Compared with English language, Textual Entailment in Arabic language faces various challenges due to the features of Arabic language. One of these challenges is lexical ambiguity, which is the difficulty to process texts with missing diacritics. Table 1 is an example of lexical ambiguity in Arabic language, where the root Alema (علم) has seven surface forms in the presence of diacritics marks [5]. The table shows the word in Arabic, its transliteration according to BuckWalter transliteration [6], each word Part of speech and its meaning in English.

The second challenge is the syntactic flexibility of the Arabic sentences [5], which is shown in the following example. The two sentences in this example have the same meaning, even they ordered differently.

عصفورٌ فوق الشجرة
فوق الشجرةِ عصفور

Another challenge of Arabic language is its richness in synonyms, where more than one word surface may have the same meaning. Table 2 is an example of different words with the same meaning [7].

Table 1: Lexical Ambiguity in Arabic Language

Arabic form	Transliteration	PoS	English meaning
عِلْمٌ	Elmٌ	N	Knowledge
عَلَمٌ	Alَmٌmٌ	N	Flag
عَلِمَ	Alِmَ	V	Knew
عُلِمَ	Olِmٌَُ	V	Is known
عَلَّمَ	Alَّmٌَُ	V	Taught
عَلِّمْ	Alَّmٌْ	V	Teach!
عُلِّمَ	Olِّmَُ	N	Is taught

Table 2 :Word Synonyms Examples in Arabic Language

Word type	Arabic word	English meaning	Synonyms
Noun	حجاب	Veil	حاجِز ، ساتِر ، فاصِل، مانِع، مُغَطٍّ، حاجِز ، ساتِر ، سَدّ، فاصِل، مُغَطٍّ
Verbs	بدأ	To begin	أَخَذَ , أَنْشَأَ , إِبْتَدَأَ , إِسْتَهَلَّ , إِفْتَتَحَ , باشَرَ بِـ , باشَرَ , باشَرَ بِـ , تَوَلَّدَ مِنْ , جَعَلَ , دَشَّنَ , شَرَعَ , شَنَّ , صَدَرَ , صَدَّرَ , طَفِقَ , عَلِقَ , فَتَحَ , قَامَ , نَتَجَ , نَشَأَ , نَمَا , هَبَّ , هَمَّ بِـ , أَشْرَقَ , إِتَّضَحَ , إِسْتَبَانَ , إِنْجَلَى , إِنْكَشَفَ , انْكَشَفَ , بانَ , بَانَ , بَرَزَ , بَزَغَ , بَلَجَ , تَبَدَّى , تَبَيَّنَ , تَجَلَّى , تَرَاءَى , تَكَشَّفَ , شاعَ , صَدَرَ, طَلَعَ , ظَهَرَ , عَلا , عَلَنَ , لأَحَ , لاحَ , نَتَأَ , نَجَمَ , هَلَّ , وَضَحَ , وَضُحَ
Adjective	ثابت	Firm	أَكِيد , أَبَدِيّ , أَزَلِيّ , باقٍ , خالِد , دائِم , سَرْمَد , قَاطِع , لازِب , لازِم , مُؤَكَّد , مُسْتَمِرّ , يَقِين ، أَبَدِيّ , بَاسِل , رَاسِخ , رَتِيب , سَرْمَد , شَدِيد , شُجَاع , صُلْب , فَاتِك , مَتِين , مَكِين , مُتَأَصِّل , مُتَجَذِّر , مُتَمَاسِك , مُتَمَكِّن , مُحْكَم , مُسْتَحْكَم , مِصْدَام , وَثِيق , وَطِيد ، مُسْتَقِر ، مُقِيم ، حَازِم , رَاجِح ، رَصِين ، مُتَفَوِّق ، مُتَمَكِّن , رَاكِد , رَزِين ، رصِين

The proposed Textual Entailment system is based on using lexical and semantic matching between the Text and the Hypothesis. The lexical matching includes two steps: the first step is computing words overlap between the pairs of Text-Hypothesis. This step includes a semantic analysis approach that is based on synonyms matching. The second step is to apply the bigram match between both of Text and Hypothesis. According to the result of the two steps, the system will build its decision on each pair relation, whether it is "Entails" or "Not Entails". The system is tested on the Arabic Textual Entailment Dataset [8].

Table 3 :Examples of Text-Hypothesis pairs

Text	Hypothesis	Judgment
بدأت قمة مجموعة الدول الصناعية الثماني الكبار أعمالها الخميس في منتجع دافوليه الساحلي في فرنسا وسط توقعات بأن تسيطر الانتفاضات التي تشهدها عدد من الدول العربية على اهتمامها	قمة مجموعة الثماني تبدأ أعمالها بمشاركة مصر وتونس	NotEntails
محكمة جنايات شمال القاهرة قضت اليوم الخميس بالسجن المشدد 5 سنوات على وزير الإسكان السابق أحمد المغربي وعزله من وظيفته وتغريمه 72 مليون جنيه مصري إثر إدانته بإهدار المال العام والتربح من وظيفته	الحكم ب السجن المشدد 5 سنوات على وزير الإسكان المصري السابق أحمد المغربي	Entails

The dataset has been annotated for entailment judgment, either Entails or Not Entails [9]. Table 3, represents examples of Textual Entailment in Arabic Language from this dataset.

The evaluation of the system is measured based on the precision and recall of the system on the inference judgment.

The rest of this paper is organized as follow: section two represents a literature review of current Textual Entailment techniques in both Arabic and English languages. Section three details the issues of Arabic language that make the Textual Entailment more challenging task. Section four discuss the used dataset for testing the system. Section five represents the architecture of the proposed system. In section six, results and evaluation are represented and finally the conclusion is discussed in section seven.

2. Related Work

Textual Entailment is defined by saying that the ability to infer if a sentence "Hypothesis" true or not by reading another sentence "Text" [10]. Textual entailment gained interest in Natural Language Processing (NLP) community as it could be adopted in different applications that apply NLP.

2.1 *The RTE Challenges:*

In 2005, Recognizing Textual Entailment (RTE) challenges have been organized for Textual Entailment systems competition. The participated systems must deal with pairs of various entailment reasoning levels (morphological, lexical, syntactic and Semantic) and pairs for different applications. The RTE challenges are considered the main source for

Textual Entailment techniques, applications and datasets [11].
Table 4 summarize the growth of Textual Entailment dataset within the RTE challenges. The dataset of the challenges have been studied as it considered a target for different Textual Entailment application and research.
The targeted applications in the first four challenges (RTE-1, RTE-2, RTE-3 and RTE-4) were Information Extraction, Information Retrieval, Summarization (SUM) and Question Answering (QA). The last three challenges were all about Summarization [12].

Table 4 :RTE-Challenges Datasets

Challenge	Dataset
RTE-1	The used dataset in the first challenge contained manually collected pairs of Text-Hypothesis. Text (T) length was 1-2 sentences and the Hypothesis (H) length was 1-2 sentence [13] [12].
RTE-2	The dataset in this challenge contained 800 pairs of text-hypothesis, the set was splitted into training set and testing set, and then the dataset has been divided into 200 text-hypothesis pairs for different applications [2].
RTE-3	No major changes on the dataset [11].
RTE-4	The used dataset was similar to RTE-3. The different was that the dataset contained 1000 pairs of text-hypothesis instead of 800 and the "Text" was longer than the text in RTE-3 where the "Hypothesis" length was the same [11].
RTE-5	The distinguish feature is that Textual Entailment judgment is done over a real corpus of Summarization scenario [14].
RTE-6	Saves the same features of RTE-5 [12].
RTE-7	Saves the same features of RTE-5 and RTE-6 [12].

In the first two challenges, the task was to determine the relation for every pair of Text-Hypothesis, whether its Entails or Not Entails. RTE-3 has made discrimination about the dataset and included a subtask named" "Extending the Evaluation of Inferences from Texts", have been initiated to provide more precise information to distinguish unknown entailment from entailment. The subtask was as follow: the first was to enable judgment system to provide more precise information about the pairs of text, the decision includes: "YES" entails, "No" does not entail and the third is "UNKNOWN", this means that the text does not provide enough information to decide if it entails the hypothesis or not.
In RTE-5 and RTE-6 a subtask was added to the previous task, where participated system should find every sentence in a document that entails a specific hypothesis. RTE-7 added a new subtask of finding new information within documents.

The used techniques for Textual Entailment in the RTE challenges were WordNet, word Knowledge, word overlap, logical inference, statistical lexical relation, and syntactic matching. Classification using Machine Learning (ML), PP, web-based statistics, N-gram and syntactic matching including entailment corpus acquisition and background Knowledge.

The accuracy (correctly judgment of entailment) was the main measure to evaluate the participated systems in RTE-2. The second task was to measure the confidence of entailment judgment, which has been measured by calculating precision of results. Those two measures have been used by many Textual Entailment applications to test their efficiency.

2.2 *Related Work in Arabic Language*

There are few number of published researches in Textual Entailment for Arabic language, which has motivated to work on this topic.
The related work in Arabic language are either limited to use the same technique applied in case of English language [5] or they handle a very specific feature such as using Negation [15].
M. Alabbas in her paper [5] was the first to target Textual Entailment issue for Arabic language, the main focus of the research was to highlight the problems imposed by Arabic language regarding RTE. Also, the authors researched the applicability of existed Textual Entailment systems for the Arabic Language. The technique the authors used was matching Text-Hypothesis pairs using Tree Edit Distance (TED) algorithm. The results of the research were as follow:

1. Combining the output of multiple data-driven dependency parsers can produce more accurate results.
2. Combining the output of three different taggers can also produce more accurate results than either parser produces by itself.
3. These experiments of the research included that obtaining dependency trees from Arabic text is an inherently difficult task [5].

F. Al-khawaldeh focused in her paper [15] on the importance of contradiction classification in RTE systems, which must be included to provide more accurate entailment results, because contradiction might reverse the actual polarity of the sentence. Contradiction words are treated as Stop words and eliminated from the text in the preprocessing stage before the entailment judgment is done. The following example shows how eliminating such words will cause problems:

T: أنا أحب قراءة الكتب
H: انا لا احب قراءة الكتب

Removing the contradiction word Laa (لا), if removed will cause wrong result as it will be considered ENTAIL. Solving the contradiction problem in Arabic language helped in improving the results accuracy in

Arabic TE systems, by marking the pairs: Negative, Positive or Neutral [15].

Bakari et al. in their paper [16] deal with the automatic Arabic text comprehension of question answering. The goal was to understand a given text then answer a list of questions related to it. To do that, the research have proposed an approach with which can analyze a given text in an open domain and generate from them logical representations. Our approach is based on recognizing the Textual Entailment method [16].

2.3 Related Work in English Language:

A lot of works were published in English language for Textual Entailment. These works cover different resources, such as: WordNet, Lexical and Semantic Matching.

Feng et al. [17] in their research have proposed a new method for lexical entailment measure, which is based on exploiting the information in the WordNet glosses, without dependence of the corpus [17].
Du et al. [18] uses lexical and semantic matching for Textual Entailment judgment. They adopt the word overlap technique at first for detection the entailment relation, and then the semantic information is also combined for this task by the use of WordNet which is a structured lexical knowledge base. The part-of-speech and named entity could give useful indication for entailment recognition and give additional attention to them [18].
Majumdar et al. [19] have developed a lexical based Textual Entailment system. The developed system is a simple lexical system that detects entailment based on word overlap between the Text and Hypothesis. The system is mainly designed to incorporate various kind of co-reference that occurs within a document and how they take an active part in the event of Text Entailment [19].

Pakray et al. [20] proposed a hybrid system using lexical and syntactic features. A two-way Textual Entailment recognition system that uses lexical and syntactic features has been developed. The hybrid TE system is based on the Support Vector Machine that uses twenty three features for lexical similarity and the output tag from a rule based syntactic two-way TE system as another feature. The important lexical features that are used in the present system are: WordNet based unigram match, bigram match, stemming, longest common subsequence, skip-gram, named entity matching and lexical distance. In the syntactic TE system, the important features used are: subject-subject comparison, subject-verb comparison, object-verb comparison and cross subject-verb comparison. The hybrid system has been developed using the collection of RTE-2 test annotated set, RTE-3 development set and RTE-3 test gold set. Evaluation scores obtained on the RTE-4 test set (includes 1000 text-hypothesis pairs) shows 55.30% precision and 58.40% recall for YES decisions and 55.93% precision and 52.80% recall for NO decisions [20].
The motivation to work on this research is the prominence of Textual Entailment for many applications in NLP, specifically in Arabic language. Available solutions have limited results and there still many issues to consider.

3. Challenges of Textual Entailment in Arabic Language

3.1 Lexical ambiguity:

The Arabic language has a productive derivational morphology, where from a single root many forms could be derived. These derived words become confusable in case diacritics are missing [5]. The problem in ignoring the lexical role of the word in the sentence is that two different words (have the same form) will be marked as identical. Also, short vowels and some phonetically distinctive items could be omitted.

3.2 Syntactic ambiguity

The structural analysis of Arabic sentences could be a complicated task. First, Arabic language has a flexible syntactic order in its sentences, where the Verb (V), Subject (S), Object (O) might come in any of the following orders, under specific conditions: VSO, SVO, OVS and VOS. Also, in Arabic sentences the subject might be absent, which also cause problems as the verbs in Arabic might be intransitively or transitively [5].

4. The Used Data Set

The used dataset in this research is ArbTEDS, which was built by Alabbas [8], is available online with open access [21]. The dataset contains 618 pairs of Text-Hypothesis in various subjects, such as: sport, politics, business and general news. To build the corpus, the dataset has been collected automatically by writing queries to google using standard Google-API and extracting text expression that entail or do not entail the query.
The collected Hypothesis in the corpus are headlines from Arabic websites of TV channels and newspapers, where the text is a sentence from the article itself. In order to ensure more accurate results and more reliable dataset for both training and testing the following steps have been done:

- In order to avoid getting similar sentences, common words number between a Text and its corresponding Hypothesis should not exceed a specific threshold.
- Small headlines are avoided by specifying the headline length is more than a predefined number of words [5].

- Headlines from one source have been matched with sentences from another source from the same story.
- The Text is not chosen from the leading paragraph to avoid bias in the collected dataset [8].

After collecting the pairs it annotated manually using an online annotation system, and to assure accuracy each pair is annotated by three different annotators. Unlike RTE challenges, this dataset wasn't built for specific application. The dataset is available in BuckWalter transliteration form.
For experiments, 100 pairs of the dataset were selected randomly in order to test the system accuracy. For the purpose of evaluating the proposed approach, the selected sentences were preprocessed before applying the proposed Textual Entailment system. The preprocessing included the following:

1. Since the dataset was represented by transliteration and in order to increase the matching accuracy, reversed transliteration were conducted to present the words using Arabic alphabets.
2. Arabic language is a very rich language in morphology. So in order to improve the accuracy of system matching results, selected sentences were processed to extract their stems. For example, the singular and plural of the same word were transformed into their stem. The following pair represent an example where stem matching improve the overlap matching accuracy. The two words "ســـاقيه" and "الســـاقين" are the same, and without applying stemming they will not be matched.

Text: " يقول الأطباء في الولايات المتحدة إن أحد المصابين بالشلل تمكن من استعادة جزء الحركة في ساقيه بفضل عملية غير مسبوقة و ذلك باستخدام التيار الكهربائي"

Hypothesis: " استخدام التيار الكهربائي في علاج شلل الساقين"

5. Textual Entailment Method

The Textual Entailment method, which has been built using Python language, is based on lexical and semantic matching. The method includes two phases: the first phase is to calculate words overlap between the Text and Hypothesis (number of similar words between them). The word overlap should be more than a specific threshold that was determined based on experiments. Besides that words which are synonyms are considered similar and increase the word overlap value. The second phase is to calculate the bigram match between each pair of Text-Hypothesis, which also should be beyond a predefined threshold value.

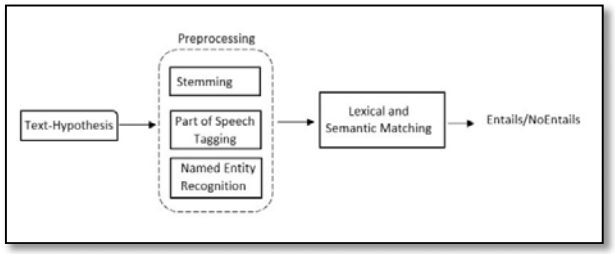

Figure 1: System Architecture

The detection system takes a pair of text-hypothesis as an input and the result of those both calculations, the system builds its decision, whether it is Entails or Not Entails.
The components of the proposed system is shown in figure 1. After preprocessing, the pairs are processed by lexical matching component, this component will produce the final entailment decision, which is Entails or NotEntails.

The preprocessing of the input text consists of three steps: The first is the Name Entity Recognition, where name entities were specified in the dataset.
Part of Speech (PoS) Tagging, which assigns a syntactic role for each token in the sentence. The used dataset has PoS assigned using both of MSTParser and Mada combined together, where both obtained an accuracy level of 81% after testing using the Penn Arabic TreeBank and recorded using CoNLL format.
Finally, in the Stemming, the stem of each token is produced by removing prefixes, suffixes and infixes. The ISRI stemmer was used [22], which could be added as a library to Python programming language.
After that, the pairs are processed in the lexical and semantic matching system, which includes the following:

Removing Stop words: Stop words have been removed before calculating word overlap, which produced more accuracy results. However, according to [15], stop words could play an important role in entailment decision, which might be incorporated in the system in future work.

Calculating Word Overlap: The word overlap is calculated between the Text and the Hypothesis to provide the basic entailment decision. It calculates number of similar words appear in the Text and Hypothesis together, where word overlap must be larger than a specific threshold. The best threshold value has been determined according to experiments. The following example shows a high overlap value between the Text and the Hypothesis:

Text: " بعد 25 عاما على الهواء, بثت اليوم آخر حلقات برنـامج أوبرا وينفري, صـــاحبـة البرنـامج الحواري و الاجتماعي ذائع الصيت"

Hypothesis: " بث آخر حلقة من حلقات برنامج اوبرا و ينفري"

The semantic matching has been incorporated in the word overlap step, that if two words are synonyms this will increase the overlap value. The next pair is an example where synonyms matching increase the overlap value. The two words "ابن" and "نجل" are synonyms.

Text: " كيم جونغ اون, الابن الاصـــغر للزعيم الكوري الشـــمالي كيم جونغ ايل, يبدأ زيارة للصـــين في خطوة ينظر اليها المحللون كدليل آخر على اختياره خلفا لوالده"
Hypothesis: "نجل الزعيم الكوري الشمالي يزور الصين"

Checking Bigram Matching: the final step in lexical matching was to apply the bigram match for entailment decision. The Hypothesis bigrams were matched with the text. In order to compute the bigram match, the measurement from [18] has been used, which is number of matched bigrams in the Hypothesis divided by number of all bigrams in the hypothesis.

Bigram Match = (Number of matched bigrams in both text and hypothesis /Number of bigrams in the hypothesis).

As in word overlap, the bigram match value should be larger than a specific threshold to determine the inference relation. The best threshold value has been determined according to experiments as well. Next pairs have a high value in bigram matching, where the following bigram expressions exist in both Text and Hypothesis: "السجن المشدد", "المشدد 5", "5 سنوات", "سنوات 5" and "أحمد السـابق", "وزير الإسـكان", "على وزير", "المغربي".

Text: " محكمة جنايات شـمال القاهرة قضـت اليوم الخميس بالسـجن المشدد 5 سنوات على وزير الإسكان السابق أحمد المغربي و عزله من وظيفته و تغريمه 72 مليون جنيه مصـري إثر إدانته بإهدار المال العام و التربح من وظيفته"

Hypothesis: " الحكم بالسجن المشدد 5 سنوات على وزير الإسكان المصري السابق أحمد المغربي"

6. System Evaluation

The system, which was built using Python language has been evaluated using precision and recall (equation 1 and 2 respectively), both are the main evaluating measures in the second RTE challenge. The developed system has been compared with the human judgment, where each pairs is annotated manually by three different annotators.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

Evaluation scores that obtained by testing the system are as follow (shown in table 1 and 2): applying the system over 100 pairs (selected randomly from the dataset) result in precision and recall of 61%, 60% respectively for Entailst judgment. NotEntails Judgment scored 58%, 59% for both precision and recall. The overall precision and recall of the system is 59%.

Table 5: Number of Entailment Judgment

Entailment Judgment	Number of entailment in the dataset	Number of correct entailment in the system	Number of given entailment by the system
Entails	42	27	40
NotEntails	58	35	60
Overall	100	61	100

Table 6: Precision and Recall that scored by the proposed system

Entailment Judgment	precision	recall
Entails	0.68%	0.64%
NotEntails	0.58%	0.60%
Overall	0.61%	0.61%

7. Conclusion and future work

Textual Entailment is recent research field that need further research and improvements, specifically in Arabic language which impose more challenging issues than other languages as English. This paper has proposed a basic lexical and semantic matching based on word overlap, synonyms match and bigram matching for finding Textual Entailment inference in Arabic language. The system result in 61% precision and recall on the selected pairs.
The proposed system shows good entailment results for Arabic language. However, these scores could be improved by testing the applicability of the following measure: considering the PoS of word in the sentence as a discriminate value or giving higher scores for name entities, where [18] concluded that Name Entity, noun and verb play more important role in determining the inference relation between the text and the hypothesis, so they could be given a higher score when word overlap is calculated, which expected to result in better accuracy.

8. References

[1] I. Dagan and O. Glickman , "Probabilistic Tex-tual Entailment: Generic Applied Modeling of Language Variability," PASCAL workshop on Learning Methods

for Text Understanding and Mining Understanding and Mining, 26 – 29, 2004.

[2] R. Bar-Haim, I. Dagan , B. Dolan, L. Ferro, D. Giampiccolo, B. Magnini and I. Szpektor, "The Second PASCAL Recognising Textual Entailment Challenge," In Proceeding of second PASCAL chalenge workshop on recognizing textual entailment, , Italy, 2006.

[3] P. Pakray, S. Bandyopadhyay and A. Gelbukh, "Textual entailment using lexical and syntactic similarity," *International Journal of Artificial Intelligence & Applications (IJAIA),* vol. 2, no. 1, pp. 43-58, 2011.

[4] A. Lavelli and R. , "aclweb.org," 29 June 2016. [Online]. Available: https://aclweb.org/aclwiki/index.php?title=Textual_Entailment_References. [Accessed 20 July 2016].

[5] M. Alabbas, "ArbTE: Arabic Textual Entailment," *in Proceedings of the 2nd Student Research Workshop associated with RANLP,* pp. 48-53, 2011.

[6] N. Habash, A. Soudi and T. Buckwalter, "On Arabic Transliteration," in *Arabic Computational Morphology*, Springer, 2007, pp. 15-22.

[7] A. Awajan, "Semantic Similarity Based Approach for Reducing Arabic Texts Dimensionality," *International Journal of Speech Technology,* vol. 19, no. 2, pp. 191-201, 2015.

[8] M. Alabbas, "A Dataset for Arabic Textual Entailment," *Proceedings of the Student Research Workshop associated with RANLP,* pp. 7-13, 2013,.

[9] M. Alabbas and A. Ramsay, "Natural Language Inference for Arabic Using Extended Tree Edit Distance with Subtrees," *Journal of Artificial Intelligence Research 48,* pp. 1-22, 2013.

[10] D. Majumdar and P. Bhattacharyya, "Lexical Based Text Entailment System for Main Task of RTE6," *CiteSeer.*

[11] D. Giampiccolo, H. T. Dang, B. Magnini, I. Dagan, E. Cabrio and B. Dolan, "The Fourth PASCAL Recognizing Textual Entailment Challenge," *Proceedings of TAC 2008.,* 2008.

[12] L. Bentivogli, P. Clark, I. Dagan and D. Giampiccolo, "The Seventh PASCAL Recognizing Textual Entailment Challenge," Proceedings of TAC, 2011.

[13] I. Dagan, O. Glickman and B. Magnini, "The PASCAL Recognising Textual Entailment Challenge," *Springer,* vol. 3944 , pp. 177-190, 2006.

[14] L. Bentivogli , I. Dagan , H. T. Dang, D. Giampiccolo and B. Magnini, "The Fifth PASCAL Recognizing Textual Entailment Challenge," *In Proc Text Analysis Conference, CiteSeer,* 2009.

[15] F. T. AL-Khawaldeh, "A Study of the Effect of Resolving Negation and Sentiment Analysis in Recognizing Text Entailment for Arabic," *World of Computer Science and Information Technology Journal (WCSIT),* vol. 5, no. 7, pp. 124-128, 2015.

[16] W. Bakari, O. Trigui And M. Neji, "Logic-based approach for improving Arabic question answering," *IEEE International Conference on Computational Intelligence and Computing Research,* pp. 1 - 6, 2014.

[17] J. Feng, Y. Zhou and T. Martin, "Recognizing Textual Entailment based on WordNet," *Second International Symposium on Intelligent Information Technology Application,* pp. 27 - 31, 2008.

[18] Y. Du , C. Yao and J. Liu, "Recognize Textual Entailment by the lexical and semantic matching," *Computer Application and System Modeling (ICCASM), 2010 International Conference,* vol. 2, pp. V2-500 - V2-504, 2010.

[19] D. Majumdar and P. Bhattacharyya, "Lexical Based Text Entailment System for Main Task of RTE6," *CiteSeer,* 2010.

[20] P. Pakray, S. Bandyopadhyay and A. Gelbukh, "A Hybrid Textual Entailment System using Lexical and Syntactic Features," *Cognitive Informatics (ICCI), 2010 9th IEEE International Conference,* pp. 291-296, 2010.

[21] M. Alabbas and A. Ramsay, "Arabic Textual Entailment Dataset," [Online]. Available: http://www.cs.man.ac.uk/~ramsay/ArabicTE/. [Accessed May 2016].

[22] K. Taghva, R. Elkhoury and J. S. Coombs, "Arabic stemming without a root dictionary," *Information Technology: Coding and Computing, 2005. ITCC 2005,* vol. 1, pp. 152-157, May 2005.

Mariam S. Khader She is a PhD Candidate in computer science at Princess Sumaya University for Technology (PSUT), Amman, Jordan. She received the BSc degree in computer networking systems from the World Islamic Science & Education University (WISE) in 2012, Amman, Jordan. She received her MSc Degree is in IT security and digital criminology in 2014 from PSUT. Between 2012-2015, she was teacher assistant and then a lecturer at the network department in WISE University.

Dr. Arafat Awajan is a full professor at Princess Sumaya University for Technology (PSUT). He received his PhD degree in computer science from the University of Franche-Comte, France in 1987. He held different academic positions at the Royal Scientific Society and Princess Sumaya University for Technology. He was appointed as the chair of the Computer Science Department (2000-2003) and the chair of the Computer Graphics and Animation Department (2005-2006) at PSUT. He had been the dean of the King Hussein School for Information Technology from 2004 to 2007, the Dean of Student Affairs from 2011- 2014 and the director of the Information Technology Center in the Royal Scientific Society from 2008-2010. He is currently the dean of the King Hussein School for computing Sciences. His research interests include: Natural Language Processing, Arabic Text Mining and Digital Image Processing.

Dr. Akram Alkouz is an assistant professor at Princess Sumaya University for Technology since 2013. Alkouz got his Ph.D. from Berlin Institute of Technology. He has research interest in NLP, Big Data Science, and Machine Learning.