# Paraphrasing Identification Techniques in English and Arabic Texts

Marwah Alian
*Hashemite University*
*Princess Sumaya University for Technology*
marwah2001@yahoo.com

Arafat Awajan
*Princess Sumaya University for Technology*
awajan@psut.edu.jo
Amman, Jordan

2020 11th International Conference on Information and Communication Systems (ICICS) 978-1-7281-6227-0/20/$31.00 ©2020 IEEE 10.1109/ICICS49469.2020.239485

***Abstract*— Rewriting sentences using different words leading to the same meaning of the original sentence is called paraphrasing while paraphrasing identification task is the task of detecting the sentence that is paraphrased from another. This research provides a literature survey of researches that have studied and proposed methods for paraphrasing identification of English and Arabic texts. Accordingly, the impact of paraphrasing on NLP applications and paraphrasing benchmarks are also studied. The comparative study on the available paraphrasing identification techniques shows that the best precision is provided with WordNet based techniques while the best accuracy is provided by deep learning with statistical features of English texts. In the case of Arabic, the best precision is provided by the distributed word vector representations with Convolutional Neural Network.**

***Keywords—* *Paraphrasing identification, Arabic text, English text, paraphrasing benchmark.***

## I. Introduction

Paraphrasing identification is the restatement of a sentence in order to produce a new sentence with the same meaning using other words and structures [1]. Paraphrasing task together with semantic similarity is used in many natural language processing (NLP) applications such as Question-Answer, Information Retrieval, text entailment, Plagiarism detection and others [2] to improve their performance and accuracy.

The semantic similarity has an important role in paraphrasing detection since the similarity score between texts is computed and used for identifying whether they are similar or not [3]. The approaches used for semantic similarity between sentences can be categorized according to Alian and Awajan [4] into four categories: co-occurrence based approach, statistical corpus based approach, feature based approach and word embedding based approach. The co-occurrence based approach represents texts as a bag of words vector while the statistical corpus based approach uses Latent Semantic Analysis which leads to representing texts as vectors in the reduced dimension space. The feature based approach takes into consideration words similarity and order similarity between texts where alignment between words that have the same Part of Speech type is performed while the WordNet based measures are used computing the semantic similarity between words and the overlapping between words order in the two texts is measured. The word embedding approach has the property of taking the word context into consideration when representing words in the distributed space [4].

However, the significant work on paraphrasing identification of English text is conducted and published while it is still limited for Arabic texts. English work has a performance reached 0.815 in terms of precision [5] while Arabic work performance achieves 0.88 [6] in terms of precision.

The research of paraphrasing identification task need to be implemented and tested using significant and well-built dataset. However, Some researchers constructed their own datasets that are used in Paraphrasing identification task while others use a common benchmark that is released and available for the research community such as Microsoft Research Paraphrase Corpus and SemEval-2015 Twitter Paraphrase Corpus. More benchmarks are given in Section 3.

The objective of our research is to analyze previous work on Paraphrase identification of English and Arabic texts and to compare their performance to identify where new research can be conducted.

This study is divided into six sections: Section II discusses paraphrasing background which includes: Paraphrasing definition, benchmarks and evaluation metrics that are used in the evaluation of researchers' proposed work related to paraphrasing. Sections III and IV provide a review for paraphrasing identification in English and Arabic texts respectively while section V introduces the impact of paraphrasing identification on other NLP applications. Finally, Section VI is the conclusion.

## II. Paraphrasing Background

### *A. Paraphrasing Definition*

Paraphrasing is the use of alternative words or expressions while maintaining similar meaning of the main sentence. If we have two texts A and B, then they are "paraphrases" of one another if A and B have the same meaning [7]. There are different types of paraphrasing such as transformational, attenuated, lexical, derivational, and real-world. The transformational paraphrasing is based on transformational relationships such as: declarative: yes/no question), extraposition, nonextraposition, active, passive, determiner - relative clause and adverb- final or not final [7]. For example, let sentence A be: "the dog bit the man" then using active, passive voice and determiner - relative clause transformational relationships to produce a new transformed sentence B which is written in Table I.

TABLE I: Applying transformational relationships for English sentences

| Sentence (A) | Transformation Rule | Transformed sentence (B) |
|---|---|---|
| The dog bit the man | Active, passive | The man was bitten by the dog. |
| The dog bit the afraid man | Determiner, relative clause | The dog bit the man who was afraid |

In this example the sentence "The dog bit the man" which is an active sentence is transformed to the passive sentence "The man was bitten by the dog" while for the second sentence "The dog bit the afraid man" we use the relative clause "who was afraid" instead of the determiner "afraid" to produce the transformed sentence "the dog bit the man who was afraid".

In Arabic language, paraphrasing refers to sentences which are considered paraphrases of each other if the first sentence shares the same words with the second sentence (i.e. they are identical) except one word that is a synonym of the word in the second sentence. Also, a set of rules for transforming one sentence to another one are used to getting paraphrases. These rules are called transformation rules [7] such as permutation, addition, deletion, reduction, expansion and replacement. However, these transformation rules do not necessarily produce paraphrases all the time. The transformed sentence may have a semantic relation with the original sentence such as entailment or it may provide a different meaning.

For example; Let sentence A be : "قرأ محمد الدرس qrA mHmd Aldrs" then permutation and replacement rules are applied to this sentence to get a new transformed sentence (B) as shown in Table II.

TABLE II: TRANSFORMING ARABIC SENTENCES EXAMPLE

| Sentence (A) | Transformation Rule | Transformed sentence (B) | Paraphrased or not |
|---|---|---|---|
| قرأ محمد الدرس<br>qrA mHmd Aldrs | Permutation | محمد قرأ الدرس<br>mHmd qrA Aldrs | Paraphrased |
| قرأ محمد الدرس<br>qrA mHmd Aldrs | Replacement | قرأ محمد الكتاب<br>qrA mHmd AlktAb | Not paraphrased |

In this example, we use the permutation rule which changes the order of the words in sentence A to get a new sentence B. Thus, the order of the words "قرأ qrA" and "محمد mHmd" is changed and we have the new sentence " محمد قرأ الدرس mHmd qrA Aldrs". While using the replacement rule which replaces one word with another one in the same position we have the word "الدرس" is replaced by the word "الكتاب AlktAb" so the new sentence is "قرأ محمد الكتاب qrA mHmd AlktAb" but in this example we do not get paraphrased sentences.

*B. Paraphrasing Benchmarks*

One of the frequently used benchmarks for evaluating approaches used in paraphrasing identification and semantic similarity measurement of English texts is the Microsoft Research Paraphrase Corpus [8] which is an English corpus built by Microsoft team which contains 5800 sentence pairs. These sentences are extracted from web news then only one sentence from each news article is selected. Thus, information about sentences is also provided such as the author and the source of each sentence. The sentence pairs are annotated by humans to indicate if a pair of sentences is paraphrased or if it has semantically similar meaning.

Another benchmark is the Twitter Paraphrase Corpus (PIT-2015) introduced by [9] and used in the SemEval-2015 task on Paraphrase and Semantic Similarity in Twitter (PIT). It consists of two sets with paraphrase annotations, training and testing. The training set contains 17,790 sentence pairs and the testing set consists of 972 sentence pairs. This dataset includes sentences that represent the real usage of informal language which has sentences that are similar to their lexical form, but they have different meanings. Thus, lexically diverse paraphrases are also included.

For Modern Standard Arabic (MSA) there is a new dataset built using a neural network which is introduced by a team form Carnegie Mellon University [10]. It can be described as a large Arabic paraphrase dataset that contains over 88K phrase pairs. The authors present and compare two methods for sentence paraphrasing in Arabic: phrase-based and neural method where the former provided good results when utilized for constructing paraphrase sentences while the latter whose results are still not as good as the former approach but the model introduces a deep learning based approach for Arabic Linguistic constructs and phrases such as those used at the beginning of a sentence.

This motivates the authors to construct a new Arabic paraphrasing benchmark to be used by researchers for the paraphrasing identification task and measuring semantic similarity between Arabic short texts [11]. However, this benchmark has been constructed by experts of Arabic language based on a set of transformation rules inspired from Chomsky rules [12] such as permutation, replacement, addition, deletion, reduction and expansion [7]. It consists of 1011 Arabic sentence pairs where the second sentence in each pair is transformed from the first sentence and each pair has a label for being paraphrased or not. This benchmark will be released soon.

*C. Evaluation Metrics*

Precession and recall are commonly used measures in evaluating the results in Machine Learning experiments [13]. Recall is called sensitivity and it is defined as the percentage of positive items that are predicted correctly positive as in Formula 1 [14]:

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{1}$$

Where:

TP is the true positives (items that are positive and they are correctly identified positive)

FN is the false negatives (items that are incorrectly identified as negative)

(TP+FN) is the total number of actual positives.

While Precision is called confidence which is defined as the percentage of predicted positives that are real positives as in Formula 2 [14]:

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{2}$$

Where:

TP is the true positive.

FP is the false positive (items that are negative but falsely identified as positive).

(TP+FP) is the total number of positives predicted.

Also, F-measure is used to combine recall and precision measures to get a single measure of effectiveness. F-measure is also called F1 score or F-score, which is the harmonic mean of precision and recall and computed as in Formula (3) [15]:

$$F = \frac{2 \times P \times R}{P + R} \qquad (3)$$

Where:

P is the Precision and R is the Recall as defined in Formula 1 and (2) respectively.

The accuracy is measured as the fraction of the correctly classified items as in Formula 4 [16] :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \qquad (4)$$

Where TP, TN, FN and FP are as defined previously.

## III. Paraphrasing Work of English Texts

The work of Vaishnavi et al. [5] utilizes syntactical information in the input sentences to measure the similarity. The input texts are represented in terms of grammar patterns using dependency parser where a grammar pattern has three roles: subject, verb and object. Then the similarity between two patterns is computed based on the similarity of the words that occur in the patterns which are based on string matching and semantic similarity between words. The longest common subsequence (LCS) algorithm is used for string matching and five WordNet based similarity metrics are used computing the semantic similarity between words. These metrics are Wu and Palmer, Leacock and Chodorow, Resnik, Lin, Jiang and Conrath metric. The overall similarity is computed as the sum of the similarities in the grammatical patterns then it is normalized to produce the semantic similarity score in the range 0.1. They evaluate their method using Microsoft Research Paraphrase Corpus (MSR) as a dataset and they use statistical evaluation metrics such as precision, recall and F-measure. They experiment their approach using each similarity metric individually then the experiment with combining five similarity measures. The results of the proposed approach with combining different similarity measures show a higher precision of 81.5 and F-score values 76.3.

Fernando and Stevenson [17] experimented six WordNet based similarity measures with the matrix similarity approach to compute the similarity between two sentences. In this approach not only are the most similar words taken into account, but all word to word similarities are measured and taken into consideration for sentence similarity in order to improve performance. This proposed method represents each sentence by a binary vector. In this vector, words that occur in a sentence are given the value of 1 and words that do not appear in the sentence are presented by 0. Then the similarity measure depends on multiplying the similarity matrix to the vector of the first sentence and the transpose of the vector of the second sentence where the similarity matrix represents the similarity between words measured by one of six WordNet based measures: Lin metric [18], Rensik metric [19], Lesk metric [20]. Wu and Palmer metric [21], Leacock and Chodorow metric [22], Jiang and Conrath metric [23]. They experiment their approach on the Microsoft Research Paraphrase Corpus and the results of matrix similarity approach show better results than the baseline approaches with all six metrics in terms of F-measure and accuracy while in terms of precision and recall it produces better results than some of the baseline approaches.

In the work of Satyapanich et al. [24] the judgment is that if two tweets are paraphrased from each other depends on how semantically similar they are. They use the textual semantic similarity system proposed by Han et al. [25] which obtains accurate semantic similarity. The first step is preprocessing of the two input sentences by removing stop words and replacing abbreviations then trigrams sets for the two sentences which are extracted. Latent Semantic Analysis (LSA) is used to compute word similarity based on word co-occurrence where they use Stanford WebBase dataset to obtain a word co-occurrence statistics and a sliding window of size (+/- 1 and +/- 4). The cosine similarity between two words is computed for words LSA vectors. The semantic similarity between every possible trigrams pair is computed using the semantic textual similarity system (UMBC). Also, two regression models are used: logistic regression and support vector regression where the logistic regression model provides F1 score of 0.697 with a precision of 0.706 and recall of 0.726 while Support Vector Regression provides F1 score of 0.691, precision of 0.707 and recall of 0.726.

While Socher et al. [26] have proposed an unsupervised approach for paraphrasing detection using recursive autoencoders which capture syntactic and semantic features. These features allow the comparison of words and phrases vectors. Also they introduce a dynamic pooling layer in order to get a fixed sized representation from the similarity matrix that has a variable size as a result of variable length sentences. The pooled representations provide enough information about sentences to be used as an input to a classifier to determine whether they are paraphrased or not. The evaluation of their approach is performed using Microsoft Paraphrasing Corpus and the results of utilizing dynamic pooling layer for standard RAE and recursive averaging achieves an accuracy of 75.5% and 75.9% respectively.

Besides these studies that are mentioned above, Agarwal et al. [16] have proposed a paraphrasing identification model that is based on deep neural network. The proposed model consists of a pair wise word similarity model and a sentence model. The word similarity model has the benefit of capturing the semantic information between words pair in the sentences while the sentence model is based on convolutional neural network (CNN) and a long short-term memory (LSTM) to provide semantic representations for sentences. The proposed approach is evaluated using two datasets: SemEval-2015 Twitter Corpus and the Microsoft Paraphrasing Corpus (MSPC). The results of the model show competitive results for paraphrase identification on MSPC dataset while it provides the best results on SemEval-2015 Twitter Corpus compared to other approaches.

A comparison between the paraphrase identification techniques for English text is presented in Table III. The comparison shows that the work of paraphrasing identification of English texts using similarity matrix with WordNet based measures provides better results in terms of precision while deep learning with statistical features introduces better results in terms of accuracy.

TABLE III: COMPARISON BETWEEN METHODS USED FOR PARAPHRASING IDENTIFICATION FOR ENGLISH TEXT

| Ref | year | Method | Dataset | Dataset size | Evaluation metric | results |
|---|---|---|---|---|---|---|
| Fernando and Stevenson [17] | 2008 | A combination of WordNet based similarity metrics with matrix similarity | Microsoft Research Paraphrase Corpus | 5800 sentence pairs | Accuracy | 74.1 |
| | | | | | Precision | 75.2 |
| | | | | | Recall | 91.3 |
| | | | | | F-measure | 82.4 |
| Socher et al. [26] | 2011 | Recursive autoencoder (RAE) with dynamic pooling layer | Microsoft Research Paraphrase Corpus | 5800 sentence pairs | Accuracy | 75.9 |
| Vaishnavi et al. [5] | 2013 | string matching and a combination of WordNet based similarity metrics | Microsoft Research Paraphrase Corpus | 5800 sentence pairs | Precision | 81.5 |
| | | | | | F-measure | 76.3 |
| Satyapanich et al. [24] | 2015 | Latent semantic analysis (LSA) , semantic textual similarity system (UMBC) logistic regression and support vector regression | SemEval-2015 Twitter Paraphrase Corpus | Training set (17,790 sentence pairs) and test set (972 sentence pairs) | Recall | 72.6 |
| | | | | | Precision | 70.6 |
| | | | | | F-measure | 69.7 |
| Agarwal et al. [16] | 2017 | Deep learning and statistical features | SemEval-2015 Twitter Paraphrase Corpus | Training set (17,790 sentence pairs) and test set (972 sentence pairs) | Precision | 76.0 |
| | | | | | Recall | 74.2 |
| | | | | | F-measure | 75.1 |
| | | | Microsoft Research Paraphrase Corpus | 5800 sentence pairs | Accuracy | 77.7 |
| | | | | | F-measure | 84.5 |

## IV. PARAPHRASING WORK OF ARABIC TEXT

Paraphrasing identification of the Arabic text is a challenging task and the existing techniques and tools to do this task are limited for several reasons such as the complexity in processing Arabic language as it has rich morphological features; ambiguity in the semantic analysis that is introduced by changing word order in Arabic sentences; and scarcity of available annotated datasets [27].

One of the proposed techniques of Arabic paraphrasing identification is the work of Mahmoud and Zrigui [6], in which a semantic similarity method for paraphrase detection in Arabic text is proposed. This method has three main phases: preprocessing, features extraction, similarity computation and paraphrase detection. Preprocessing phase consists of segmentation, sentences identification and tokenization. The features are TF-IDF and word vector representation where TF-IDF is used for weighting words to represent the more descriptive words in a sentence while words vectors are produced using word2vec skip-gram model then these vectors are combined to obtain sentence representation. Two similarity metrics are combined to compute the overall similarity, cosine similarity and Euclidean Distance. Two sentences are judged to be paraphrases if the similarity between the sentences exceeds a specific threshold which is determined experimentally. They evaluate their approach using the 3233 text documents from the historical category of the Open Source Arabic Corpus (OSAC) by applying the approach using each similarity metric individually then by combining the two metrics. The best results obtained by combining the cosine and Euclidean distance metrics by providing a precision of 0.85 and a recall of 0.84.

A. Mahmoud et al. [27] introduce a deep learning based approach for Arabic paraphrasing identification. This approach consists of three phases: preprocessing phase, generating words vectors and sentence vectors and capturing statistical regularities phase. In the first phase, relevant information is extracted from documents while in the second phase, words representations are generated using word2vec algorithm. The final phase uses Convolutional Neural Network (CNN) which represents the semantics of texts and can learn more contextual information. CNN is used to getting the statistical regularities in the context. The proposed approach is evaluated and provides good precision.

Another approach is proposed by Al-Smadi et al. [28] who have studied the paraphrasing identification for Arabic news tweets. This approach is composed of three phases: text preprocessing; extracting features and text classification. They focus on extracting lexical, syntactic and semantic features. The researchers argue that extracting these features is important to overcome many limitations in the existing approaches for paraphrasing identification.

These features are utilized in classification phase to train Support Vector Regression (SVR) and Maximum Entropy classifiers. To evaluate this approach, a dataset is constructed and experiments are conducted where the results provided by the proposed approach are comparable to the other baseline results. It achieves F measure equals 0.872 and Pearson correlation of 0.912 using the SVR classifier.

Table IV provides a comparison between the available techniques used for paraphrase identification of Arabic texts. It is shown in Table IV that the results of using word embedding approach are promising in terms of precision.

TABLE IV: COMPARISON BETWEEN METHODS USED FOR PARAPHRASING IDENTIFICATION FOR ARABIC TEXT

| Ref | Year | Method | Dataset | Dataset Size | Evaluation metric | Results |
|---|---|---|---|---|---|---|
| Mahmoud and Zrigui [6] | 2017 | TF-IDF and word2vec cosine similarity and Euclidean Distance | The historical documents category from the Open Source Arabic Corpus OSAC | 3233 text documents | Precision | 0.85 |
| | | | | | Recall | 0.84 |
| Al-Smadi et al. [28] | 2017 | Support Vector Regression (SVR) and Maximum Entropy classifiers | SemEval-2015 Twitter Paraphrase Corpus | Training set (17,790 sentence pairs) and test set (972 sentence pairs) | F-measure | 0.872 |
| | | | | | Pearson Correlation | 0.912 |
| Mahmoud et al. [27] | 2018 | word2vec and Convolutional Neural Network | Open Source Arabic Corpus OSAC | 22,429 text documents | Precision | 0.88 |

## V. APPLICATIONS OF PARAPHRASING IN NLP TASKS

Several NLP tasks apply paraphrasing as a part or phase in the process of handling texts. For example, Question paraphrasing is used to matching a question to a set of frequent questions that are associated with answers [29]. The work of Xu et al. [30] have introduced an open question-answer system over knowledge base and has one layer of paraphrase. The search in this system is done on Open Information Extraction Knowledge Base and Freebase to which entities in the dataset are linked to. In this system there is one paraphrase layer compared to other systems that have three layers. In this paraphrasing phase, they concentrated on word level paraphrasing not on a sentence level. The evaluation is done using the Web-Question set where the results show a better performance. The proposed system obtains F1 score of 0.379 which is better than the previous systems. While the work of Culicover [31] investigates the effect of using paraphrase relationships on retrieving texts and the content information from a stored English texts. This study has focused on suitable values of the parameters; storage form, input form, matching technique and the response form. A hand .written example is given for each theoretical part of the paper but no experiment is provided.

In information retrieval, the work of Wallis [32] tries to add the semantic level and reconstruct the documents and queries by paraphrasing. Two datasets which are used to evaluation: the first one is a set of 3200 titles and abstracts from ACM .While the second dataset consists of 620 full articles from Time Magazine. The query time for the first dataset is decreased by 33.1%.

Therefore, paraphrasing can be defined as a mutual entailment relation for a text and a hypothesis [33]. For example, Maris et al. [34] have proposed a dependency-based paraphrasing method used in the process of recognizing textual entailment. A preprocessing step is to apply dependency parsing for all sentences, texts and hypotheses. Then for each text-hypothesis pair, a process of searching and aligning a suitable paraphrase for all the dependency sub-trees in the parsed text is performed using Discovering Inference Rules from Text (DIRT) paraphrases collection. To evaluate this approach, RTE3 development set where the results show that many paraphrases are substituted and a small improvement (over 1%) is obtained on Recognizing Textual Entailment (RTE) system.

## VI. CONCLUSION

This research provides a definition for paraphrasing in English and Arabic and studies the techniques used for paraphrasing identification for these both languages. Also, a comparison between the available methods used for paraphrase identification is provided for English and Arabic. It is shown from the comparison that the research in the field of paraphrase identification for Arabic texts has recently received the popularity of the research community and provides promising results in terms of using distributed word vector representations and neural network techniques but they are still limited. While the work for paraphrasing identification in English has started early and it has been growing. For identifying paraphrasing in English texts, the use of WordNet based measures provides better results in terms of precision while deep learning with statistical features introduces the best accuracy. However, the polysemy problem still has to be studied in the case of paraphrasing identification task as the word embedding and WordNet based measures techniques do not solve the problem of words with multiple senses. This requires great efforts: human and material resources to get what we aspire to.

## REFERENCES


[1] S. S. M. Fernando, "A semantic similarity approach to paraphrase detection," in *Proceedings of the 11th Annual Research Colloquium of the UK Special Interest Group for Computational Linguistics*, 2008.

[2] N. P. A. M. S. P. O. Vo, "Paraphrase Identification and Semantic Similarity in Twitter with Simple Features," in *Proceedings of SocialNLP 2015@NAACL-HLT*, 2015.

[3] S. Srivastava and S. Govilkar, "A Survey on Paraphrase Detection Techniques for Indian Regional Languages," *International Journal of Computer Applications,* vol. 163, no. 9, p. 0975 – 8887, 2017.

[4] M. A. A. Alian, "Semantic similarity approaches-Review," in *2018 International Arab Conference on Information Technology (ACIT2018)*, Werdanye, Lebanon,, 2018.

[5] V. S. M. M. R. S. Vaishnavi, "Paraphrase Identification in Short Texts using Grammar Patterns," in *2013 International Conference on Recent Trends in Information Technology (ICRTIT)*, 2013.

[6] A. Z. M. Mahmoud, "Semantic Similarity Analysis for Paraphrase Identification in Arabic Texts, The 31st Pacific Asia Conference on Language,," in *The 31st Pacific Asia Conference on Language, Information and Computation PACLIC 31*, 2017.

[7] M. AlKouli, Transformation Rules for Arabic Language ( qwAEd tHwylyAh llgAh AlErbyAh),

Jordan: Dar Al-Falah, 1999.

[8] B. B. C. Q. C. Dolan, "Microsoft Research Paraphrase Corpus," March 2, 2005.

[9] W. R. A. C.-B. C. D. W. B. J. Y. Xu, "Extracting lexically divergent paraphrases from Twitter," *Transactions of the Association for Computational Linguistics (TACL),* vol. 2, no. 1, 2014.

[10] F. B. A. L. W. Al-Raisi, "Neural Symbolic Arabic Paraphrasing with Automatic Evaluation," in *4th International Conference on Natural Language Computing (NATL 2018)*, 2018.

[11] M. A. A. A.-H. A. A. R. Alian, "Towards building Arabic paraphrasing benchmark," in *the Second International conference on Data Science, E-learning and Information Systems (DATA' 2019)*, Dubai, 2019.

[12] N. Chomsky, syntactic structure, the Hague. Paris: Mouton publishers, 1957.

[13] D. M. W. Powers, "Evaluation: From Precision,Recall and F-Factor to ROC, Informedness, Markedness & Correlation," Adelaide, Australia, 2007.

[14] T. K.M., "Precision and Recall," in *Encyclopedia of Machine Learning*, Ed. Boston, MA: Springer,, 2011.

[15] Y. Sasaki, "The truth of the F-measure," 2007.

[16] B. R. H. L. H. R. M. Agarwal, "A Deep Network Model for Paraphrase Detection in Short Text Messages," *arXiv:1712.02820v1,* 2017.

[17] S. Fernando and M. Stevenson, "A semantic similarity approach to paraphrase detection," in *Proceedings of the 11th Annual Research Colloquium of the UK Special Interest Group for Computational Linguistics*, 2008.

[18] D. Lin, "An information-theoretic definition of similarity," in *Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998)*, 1998.

[19] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in *Proceedings of IJCAI-95*, Montreal,Canada.

[20] S. Banerjee and T. Pedersen, "Extended Gloss Overlaps as a Measure of Semantic Relatedness," in *The Eighteenth International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, 2003.

[21] Z. a. P. M. Wu, "Verb semantics and lexical selection," in *the 32nd Annual Meeting of the Associations for Computational Linguistics*, 1994.

[22] C. a. C. M. Leacock, "Combining local context and WordNet sense similarity for word sense identification," in *WordNet, An Electronic Lexical Database,The MIT Press*, 1998.

[23] J. J. C. D. Jiang, "Semantic similarity based on corpus statistics and lexical taxonomy," in *Proceedings of Research in Computational Linguistics*.

[24] T. Satyapanich and H. a. F. T. Gao, "Ebiquity: Paraphrase and Semantic Similarity in Twitter using Skipgram," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015.

[25] L. F. T. M. P. J. A. Y. Y. Han, "Improving Word Similarity by Augmenting PMI with Estimates of Word Polysemy," *IEEE Trans. on Knowledge and Data Engineering,* vol. 25, no. 6, p. 1307–1322, 2013.

[26] R. H. E. H. P. J. N. A. Y. M. C. D. Socher, "Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection," in *Proceedings of the 24th International Conference on Neural Information Process (NIPS'11 )*, 2011.

[27] A. Z. A. Z. M. Mahmoud, "A Text Semantic Similarity Approach for Arabic Paraphrase Detection," in *Computational Linguistics and Intelligent Text Processing. CICLing 2017, Lecture Notes in Computer Science, vol 10762.*, Springer, 2018.

[28] M. J. Z. A.-A. M. J. Y. AL-Smadi, "Paraphrase identification and semantic text similarity analysis in Arabic news tweets using lexical, syntactic, and semantic features," *Information Processing and Management,* vol. 53, no. 3, pp. 640-652, 2017.

[29] N. Tomuro, "Interrogative Reformulation Patterns and Acquisition of Question Paraphrases," in *the second International Workshop on Paraphrasing*, 2003.

[30] Y. M.-G. P. M. Y. G. R. Xu, "Paraphrase for Open Question Answering: New Dataset and Methods," in *Proceedings of the Workshop on Human-Computer Question Answering*, 2016.

[31] P. W. Culicover, "Paraphrase Generation and Information Retrieval from Stored Text," *Mechanical Translation and Computational Linguistics,* vol. 11, no. 1 and 2, 1968.

[32] P. Wallis, "Information retrieval based on paraphrase," in *Proceedings of the First Pacific Association for Computational Linguistics Conference (PACLING )*, Vancouver, Canada, 1993.

[33] P. B. S. G. A. Pakray, "Textual Entailment Using Lexical and Syntactic Similarity," *International Journal of Artificial Intelligence & Applications (IJAIA),* vol. 2, no. 1, 2011.

[34] E. K. E. B. W. Marsi, "Dependency-based paraphrasing for recognizing textual entailment," in *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, 2007.