Information Sciences Letters

Volume 11
Issue 5 *Sep. 2022*

Article 40

2022

State-of-the-Art: Assessing Semantic Similarity in Automated Short-Answer Grading Systems

Zaira Hassan Amur
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak, 32610, Malaysia, zaira_20001009@utp.edu.my

Yew Kwang Hooi
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak, 32610, Malaysia, zaira_20001009@utp.edu.my

Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/110540

State-of-the-Art: Assessing Semantic Similarity in Automated Short-Answer Grading Systems

*Zaira Hassan Amur** *and Yew Kwang Hooi*

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak, 32610, Malaysia

Received: 21 Feb. 2022, Revised: 22 May 2022, Accepted: 24 Jun. 2022.
Published online: 1 Sep. 2022.

Abstract: The use of semantic in Natural Language Processing (NLP) has sparked the interest of academics and businesses in various fields. One such field is Automated Short-answer Grading Systems (ASAGS) for automatically evaluating responses for similarity with the expected answer. ASAGS poses semantic challenges because the responses of a topic are in the responder's own words. This study is providing an in-depth analysis of work to improve the assessment of semantic similarity between corpora in natural language in the context of ASAGS. Three popular semantic approaches are corpus-based, knowledge-based, and deep learning are used to evaluate against the conventional methods in ASAGS. Finally, the gaps in knowledge are identified and new research areas are proposed.
Keywords: Automated Short-Answer Grading System (ASAGS), Natural language processing (NLP), semantic similarity, Question Answering Systems (QAs)

1 Introduction

Semantic similarity has a significant impact on the accuracy of natural language processing systems such as Automated Short-Answer Grading systems. ASAGS analyses and matches responses with the answer by using natural language processing. Responses in natural language can be ambiguous and therefore difficult to be understood.

Answers in ASAGS can be long, up to as many as 20 words. This is a challenge for syntactic processing. Furthermore, the order and relationship of these words affect the actual meaning of the answer. Answers provided by students may not be syntactically correct. As such, do we regard answers with poor grammar and structure but correct content as true or false? Furthermore, some answers are to be evaluated using a range of marks. What criteria are used to grade these responses? These are sources of irregularities that impact grading accuracy. These factors are similar to problems faced in text similarity analysis described in [1].

Furthermore, there are different question types in ASAGS: - factoid, descriptive, short, and long questions. Different questions types affect the intent of the answer, i.e. to assert a fact, to provide a description, or to assess. This can affect the grading accuracy.

The objective of this study is two-fold: -

1. To examine the limitations of traditional and existing methods.
2. To identify gaps in knowledge for further studies.

This paper is arranged as follows:-

Section 2 describe the background of the study. Section 3 presents the in-depth analysis of text semantic similarity in which we have identified the strengths, weaknesses and applications. Section 4 explains problems and presents open issues and provides some solutions. Section 5 proposes the method for text semantic similarity. Section 6 discuss the preliminary results and the paper finally concludes in **section 7.**

2 Background of the Study

There are various traditional measures used to extract the short text similarity, like string-based measures, corpus-based measures, and knowledge or ontology-based measures. Using standard metrics, a number of studies were able to get good results. However, because of the limitations, a variety of novel approaches have been presented. Meanwhile, the growing trend of neural networks, such as the deep learning (DL) models improves the extraction technique of semantic similarity.

Many applications such as text classification, information retrieval, and sentiment analysis employ semantic similarity and achieve positive outcomes. The practice of classifying

*Corresponding author e-mail: zaira_20001009@utp.edu.my

text into ordered groupings is known as text classification, sometimes termed as text tagging or text categorization. Text classifiers assess text using Natural Language (NL) and apply pre-defined labels or classes based on its content [1, 2].

However, sentiment analysis refers to recognizing emotions. Whereas information retrieval is used to extract the meaning of the text. Information retrieval is an active research area in the field of data mining. It uses different forms of platforms to extract the data. Like documents, PDF files, tweets, search engines. ASAGS (automated Short Answer Grading Systems) is the process of assessing students' replies on exams. Most of the present ASAGS systems evaluate scores entirely based on replies. This research article includes and proposes the model that will be utilized as the baseline model and can provide state-of-the-art results in automated short grading systems ASAGS. It is important that short texts (subjective answers) must be understandable and disambiguated so that learners can find accurate information.

However, short text similarity is also related to textual entailment (TE) and paraphrasing techniques. Which is mostly used in many natural language processing tasks. These techniques differ from each other. Textual entailment uses the direct relation among text fragments by using the hypothesis techniques, whereas paraphrasing is used to recognize the same meaning of the text, both techniques work on yes or no decisions. ASAGS uses these techniques to rate the semantic relatedness among groups of words or sentences [1]. The textual semantic similarity has been proposed in 2006, where only a small amount of text was supposed to be included but after that research becomes enhanced from short to long and long to individual words [1]. It also works well in many web applications like ontology generation, keyword extraction, and entity disambiguation.

There are many traditional methods used in natural language processing and the most common are, Bag of words, Vector space model, and BM25 [2], these methods help to generate the words in a text. In NLP usually, traditional methods are poor method that can't properly detect the semantics of text at conceptual levels. Due to the limited amount of text, it is not easy to calculate semantic similarity with these methods.

Several other methods such as n-gram-based, word-based method, long-based distance so on and so forth are used to fix the real word errors to improve the accuracy for short answers [3], but some limitations that cause the problem of duplication as well as produce low accuracy [4].

Moreover, some algorithms such TPB contains high time complexity because it uses several semantic relationships. One study mentioned the relationship diagram for short texts known as knowledge graph and random walk method to improve the concepts of optimal word segmentation [5]. However, another study provides the attention neural network approach for reading the text [6] these methods distinguish the text summarization and semantic information of the text. However, Narasimhan 2018 proposed a method that provides the input transformation to set the optimal changes to the model architecture to extract the answer from relevant questions [7]. [1] Meanwhile, a study suggests a new QA system for extracting grounded and commonsense information from the text [8]. A study presented by [9] reviewed how to gain new facts about the world based on knowledge graphs. Most of these methods are task-oriented. Although these methods have good expressiveness on particular tasks, they still lacks understanding and failed to improve the accuracy of answers. Some studies focus on external knowledge to upgrade the quality of topic identification and disambiguation in short texts. However, [10] suggest that related features from LDA can also help with disambiguation. They worked to capture semantic relations between terms using the novel approach LDA, this approach helps to improve the accuracy of short text conceptualization by using context semantics. Whereas POSs like verbs, adjectives, and other attributes, can also help to identify keywords from texts. Moreover, the study also introduces the framework for short texts that detect the errors from text [10]. More specifically, the work has been divided into three subtasks to understand the short text: Text segmentation, type detection, and concept labeling. However, another study presents an approach to solving the problem of semantic similarity in test papers, with the help of density entropy. They selected the various question papers from the item bank and then applies the calculation of semantic similarity to detect the intelligent test papers from the corpus. [4].Furthermore, in a study [11], the researchers developed the algorithm to improve the performance of STSS with low time complexity. This algorithm incorporated the different WordNet-based measures to address the word pairs with specific POSs that help to improve the evaluation of semantic similarity. Moreover, one study represented the work, that focusing one semantic textual similarity (STS) of a question pair [11]. In this study, they find if two questions have the same answers, then they are semantically equivalent [1]. To compute semantic similarity for short texts is important in many areas. Many approaches have been proposed that uses linguistic analysis. These methods determine whether the words in two short texts look alike, in terms of the largest common substring (LCS) [12]. These approaches usually work for trivial cases.

Text Semantic Similarity Analysis

Table 1: Corpus-Based Measure.

Method	Strengths	Limitations	Research
Support vector machine	It works well with a huge amount of data.	It can't tell the difference between homophones and synonyms.	[13]
Latent semantic analysis	LSA can identifies the polysemy problems from the text	LSA doesn't care about the sequence of words in a sentence.	[14]
Word2Vec	Easy to train and powerful as compared to other approaches.	Word2Vec can't deal with terms that are not familiar.	[14]
LDA	It can evaluate semantic associations between short texts.	Inefficient to identify the sequence of words in a sentences.	[14,15]

Table 2: Knowledge-Based Measure.

Method	Strength	Limitations	Research
Shortest path	Help to analyze the information disseminati on. And finds the latent relationship in weighted social networks.	Ignores necessary details and unable to solve the negative edge outcomes.	[13]
Lesk	Easy to use	Inefficient to identify necessary details. High exponential complexity.	[14]
WuP	It helps to extract the synsets from WordNet taxonomies .	It doesn't take into account how semantically related the ideas are.	[14]

In text semantic analysis we have presents the limitations and strengths of corpus-based measures. However, **Table 2** Presents the knowledge-based measures. These measures have certain limitations and strengths. These methods are frequently used to extract keywords from the text. Another measure known as String-based similarity was employed in the short text at the start of the investigation. Such as cosine, jacquard, LD, Euclidian distance, LCS have been proposed to deal with short text similarity. Like corpus and ontology-based measures, string-based measures couldn't identify the sequence of words properly. However, cosine similarity is still used in a variety of other techniques, such as deep learning and other neural networks. The calculation of short text similarity using string-based similarity metrics is still challenging. In similarity computing, we need to make it possible for machines to interpret short messages better.

Traditional text extraction methods, such as string or corpus-based metrics, and ontology-based methods, are inadequate to detect the text. The corpus-based and knowledge-based measures are also known as non-Deep learning measures. Corpus-based measures are corpus-dependent, they are used to take two or more sentences from the corpus and calculate them. However, knowledge-based measures use the concept of ontologies. These are the metrics that are used to determine how similar two or more words are. In knowledge-based measurements, the notion of WordNet is frequently utilized. Moreover, nowadays machines becomes trained by using deep learning approaches. These approaches can be used as a combined approach with corpus or string-based measures.

Some deep learning techniques are neural networks. Such as,CNN, RNN, BERT, Because of their strong characteristics, these models are commonly employed in short text similarity. The accuracy is far higher than the previous techniques.

Table 3. Applications and datasets used in Traditional semantic similarity measures

Base method	**Dataset**	**Applications**	**Research**
String Based	MSRP(Microsoft paraphrase corpus)	WordNet LCS Vector-Based	[16]
Combined (String & Corpus-based)	MSRP(Microsoft paraphrase corpus)	LCS PMI-IR LSA	[17]
Corpus-based	Gigaword & DUC-2004.	Word Embedding Vector-Based	[18]

Corpus-based	OSAC	Word Embedding Vector-Based	[19]
Knowledge-based	PILOT	WordNet Vector-Based	[20]
Corpus-based	MRPC, P4PIN, STS2015	Word Embedding Vector-Based	[21]
String & knowledge-based	M&C,R&G,WS-353	WordNet LCS Structure-based	[22]
Corpus-based	ASAGSent, MSRP	Word Embedding Vector-Based	[21,23]
Corpus-based	Kaggle	WordNet LSA	[24]

The above Table 3. Summarizes the main datasets used in traditional text similarity techniques. The Microsoft paraphrase is the most frequently used corpus employed by various studies. This dataset includes 5801 sentence pairs that may be extracted from newspapers and social media sites. Other data sets, such as Kaggle and gig word, have extensively explored text semantic similarity. WordNet, LCS, and vector-based similarity metrics, on the other hand, have been widely used by three conventional methods.

Problem Analysis in Short-Answer Grading Systems (ASAGS)

1. Purpose:

To extract the most relevant answers from

The textual information of QA pairs [25]

Technique Used:

IKAAS, LSTM, CNN, TrecQA dataset

Problem analysis:

a) 100 instances predicted incorrectly.

b) Couldn't recognize the factoid answers.

c) Unable to predict the positive answers.

d) Failed to give better performance on how-many question

2. Purpose:

To enhance the model by adding different attentive features [26]

Technique Used:

BERT (Embedding with Word2VEC & Glove),

WikiQA dataset

Problem analysis:

a) Word2VEC with glove embedding lead to poor performance on the model.

b) No major improvement while adding another convolution layer.

c) Softmax functions were added for the matrix multiplication that didn't lead to greater performance.

d) 1636 features were added through logistic regression input that exceeded the training time.

e) Poor generalization on test data

3. Purpose:

Uses web pages to improve the accuracy of answers.
Technique Used:

BMQA (Stanford NER & Alchemy NER, Sentence matching), TrecQA data set [3].

Problem analysis:

a) Combined approach exceeded the training time.

b) Sentence matching produced a good score of accuracy but it detects correct and incorrect answers together that causes the noise.

c) Sentence matching ignored the features that are not properly detected by NER, hence the accuracy was affected, and decreased.

4. Purpose:

Uses language model to understand the question and answers [15]

Technique Used:

Bert (BB-bow, BB-CNN, BB-RNN)

Problem analysis:

a) Sparsity of the training data

5. Purpose:

Helps to extract the words [27]

Technique Used:

BM25, (Tf-idf factors), TrecQA dataset

Problem analysis:

a) It only has a small amount of text

b) It considered the scores of various terms that give independent evidence of similarity.

6. Purpose:

It extracts the exact verse from the Holy

Quran by using semantic similarity **[**28**]**

Technique Used:

N-gram (BPNN Backpropagation neural network),

Reference dataset

Problem analysis:

a) The question doesn't contain the network activities of EAT, which in return cause the long sentence as an answer

b) Takes longer time processing, as the questions are not limited.

c) A number of text words in ontology produces the same meaning.

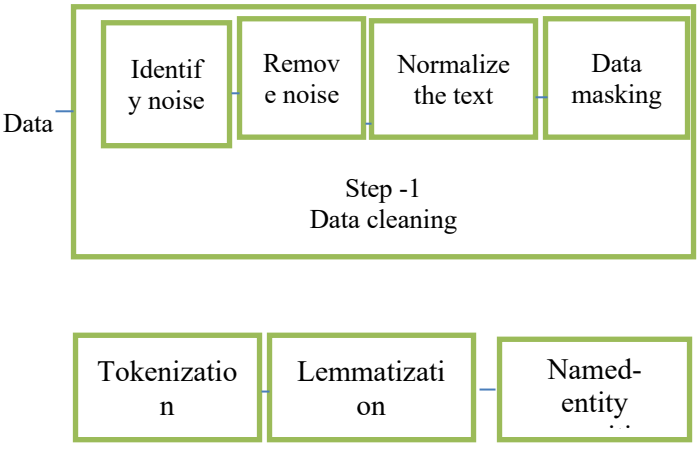
Table 4: Open Issues and Proposed Solutions.

Base method	**Dataset**	**Applications**	**Research**
String Based	MSRP(Microsoft paraphrase corpus)	WordNet LCS Vector-Based	[16]
Combined (String & Corpus-based)	MSRP(Microsoft paraphrase corpus)	LCS PMI-IR LSA	[17]
Corpus-based	Gigaword & DUC-2004.	Word Embedding Vector-Based	[18]
Corpus-based	OSAC	Word Embedding Vector-Based	[19]
Knowledge-based	PILOT	WordNet Vector-Based	[20]
Corpus-based	MRPC, P4PIN, STS2015	Word Embedding Vector-Based	[21]
String & knowledge-based	M&C,R&G, WS-353	WordNet LCS Structure-based	[22]
Corpus-based	ASAGSent, MSRP	Word Embedding Vector-Based	[21,23]
Corpus-based	Kaggle	WordNet LSA	[24]

The above table 4. Shows the issues and solutions of problems identified from several various studies. These studies show that cosine similarity can be better utilized for text extraction techniques. However, the language model Bert also contains adequate qualities such as transformers and classifiers, which can extract responses from datasets using preprocessing techniques.

Proposed Method for Short Text Semantic Similarity: Bert Language Model

Semantic similarity is still the most unresolved issue in short-answer grading systems ASAGS. There are five major categories of questions. Factoid questions, Descriptive questions, hypothetical questions, list type, and Procedural type questions. Factoid questions usually give short answers. The answer to these questions is like a sentence, a piece of a text, and requires a single answer only. This may help to clear the concept of retrieving the short answers. Table 4. Presents the research gaps from various studies. Based on the proposed solutions we believe that the major gap is to form a model that helps to retrieve the correct answers by using text classification, text summarization as well as text preprocessing techniques. Nowadays, deep learning captures higher attention in the field of semantic similarity. DL helps to improve the performance of various models through various robust features. One of the well-known and recently developed models is known as the Bert language model [29]. This model helps to understand different syntactic and semantic rules of language. In the information retrieval process, it can identify the next keyword as well as predict the next sentence. In this research study, we are proposing the model known as the Bert answer selection model [27.]The Bert Language model can understand the language and produces the text with semantic and syntactic rules [29]. Bert language model can be enhanced and produce more accurate results through the help of text classification and text preprocessing techniques. Further, we can use the cosine similarity with the Bert model to extract the most relevant answer.



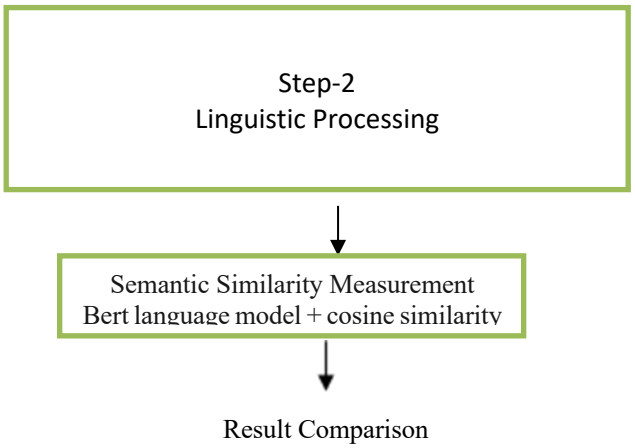


Fig. 1: Preprocessing Pipeline for automated short-answer grading systems ASAGS.

In-text retrieval systems preprocessing techniques (Fig 1.) help to reduce the semantic and syntactic problems. Data cleaning and linguistic processing are some of the fundamental rules of text preprocessing. The purpose of using data cleaning is to detect the noise from text like stop words, punctuation marks, tags, and so on. After detecting and removing the noise from text, data masking can be applied which hides the sensitive information. Whereas character normalization uses linguistic processing like lemmatize the text by removing suffixes from the text. Data cleaning and linguistic processing work simultaneously whenever they receive the input. Linguistic processing uses the tokens to identify the keyword through parts of speech tagging or named entities recognition. However, we can apply the Bert language model and cosine similarity to capture the semantics from texts. The proposed method can help to detect the most relevant answers by using data cleaning and linguistic processing techniques.

Preliminary Results and Discussion

Table 5: Shows the preliminary results of text semantic similarity.

Reference	Mean Average Precision (MAP)	Mean Reciprocal Rank (MRR)
[2]	[0.243]	[0.6775]
[3]	-	-
[25]	-	0.778
[26]	-	0.710
[27]	0.789	0.810
[29]	0.7843	0.844
[30]	0.7540	0.771

A short survey **(Table 5.)** was done to identify the performance of various models used in different studies to extract the similarity. The model proposed by [29] extracted the most relevant answers through an attention network. Lots of variations has been done in this study, after multiple alterations this model achieved good performance results. But failed to capture the short answers like factoid answers. Another study [30] uses the Bert model. After adding multiple features to enhance the model performance, this study achieves the average results. We couldn't find any major results from the study [15] but have noticed the performance of work by doing the problem analysis. The study [31] proposed the Bert model. This model has achieved better responses. But due to a lack of preprocessing techniques, some keywords couldn't be identified. Overall Bert is a novel language model that can be enhanced through various robust features. This study used various classifiers to detect the accurate answers, Like BB-Bow, BB-CNN, and BB-RNN and the results have been detected through wikiQA, TrecQA Raw, and TrecQA clean dataset. The performance of the model has affected the complexity due to the lack of preprocessing. But due to slight changes in the Bert model, the model has improved its performance on the different data sets. The results show that the Bert model has a positive impact on answer selections tasks.

The performance of the Bert language model can be best utilized by extracting the data for ASAGS. We have also noticed that the Bert model gives better results on short texts rather than long paragraphs. The study [2] enhances the BM25 model through Tf-idf factors but still needs some modifications through cosine similarity to better understand the text. We have checked that cosine similarity provides a major role in text semantic similarity. In information retrieval and related research, cosine similarity is a commonly used measure.

1. **Evaluation Metrics**

Mean average precision (MAP) and mean reciprocal rank (MRR) metrics are used to evaluate the model results **(Table 6.)**. These metrics help to extract the keywords from candidate answers and select the most relevant answer. Mean reciprocal rank is only used to rank the first suitable answer. However, mean average precision order the all matched answers present in the data set. [32].

The equations of these measures are:

$$MAP \quad (Q) = \frac{1}{} \sum_{j=1}^{|Q|} \frac{1}{} \sum_{|Q|mj \; mj_{k=1}} (Rjk) \quad [27]$$

$$MRR \quad (\mathrm{Q}) = |\frac{1}{} \sum_{j}^{|Q} Q| \quad =|_1 rj$$

[27]

Table 6: Description of parameters used in equation.

Parameters	Description
Q	Set of questions
M*j*	A number of relevant answers.
Rjk	Shows the list of candidate answers. contain the most relevant answer asked by *Qj*.
Rj	The inverse of the relevant answer rank [27]

Conclusions

The main consideration of this study is to demonstrate the several problems raised by previous studies. Based on the literature and previous work, we have also selected certain techniques, such as string, corpus, and knowledge-based measures that are frequently used in short answer grading systems ASAGS. However, there are certain drawbacks in utilizing these methods. We have also identified the key research trends and gaps from recent studies. The primary issues have been discovered through the problem analysis. Because of the semantic and syntactic restrictions, a number of models were unable to identify the short answer accurately. This research study highlights some advancements in a neural network as well. A preliminary survey has been done to check the responses from various studies.We have also noticed, if preprocessing techniques will be employed properly then there will be very limited chances of risk to appear in the text generation process. Preprocessing techniques based on tokenization and lemmatization, normalization, and text summarization. These techniques help to detect the correct answer that is most relevant to the questions. In deep learning, language models like Bert work well and can better detect the correct answers. Further, robust features like modifications in classifiers can assist to enhance the model to improve the accuracy of answers.

Acknowledgment

The work presented is supported financially by Universiti Teknologi PETRONAS.

References

[1] P. Huang, A study of using syntactic cues in short-text similarity measure.", Journal of Internet Technology, 839-850.(2019)

[2] S., Cucerzan, Improving TF and IDF factors in BM25 by using collection term frequenciesBM25-CTF: Journal of Intelligent & Fuzzy Systems, 34(5), 28872899. (2018).

[3] S. Da-Xiong, . Research on Text Error Detection and Repair Method Based on Online Learning Community. Procedia Computer Science, 154, 13-19 (2019).

[4] W. Yang, W. An Intelligent Test Paper Generation Method to Solve Semantic Similarity Problem. In IOP Conference Series: Earth and Environmental Science (Vol. 252, No. 5, p. 052126). IOP Publishing. (2019).

[5] M.Pichler, M. Analysis of word co-occurrence in human literature for supporting semantic correspondence discovery. In Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business (2014).

[6] W. Zeng. Finding main causes of elevator accidents via multidimensional association rule in edge computing environment. China communications, 14(11), 39-47 (2017).

[7] R. Narasimhan, Improving language understanding by generative pre-training. (2018).

[8] B.Wang, Commonsense for generative multi-hop question answering tasks. 1809.06309 (2018).

[9] N.Gabrilovich, A review of relational machine learning for knowledge graphs. Proceedings of the IEEE, 104(1), 11-33 (2015).

[10] H.kim . Dependent conceptualization, in Proceedings of the Twenty- Third International Joint Conference on Artificial Intelligence, (2013).

[11] H.Wang, Short text understanding through lexical semantic analysis. In 2015 IEEE 31st International Conference on Data Engineering (pp. 495-506). IEEE (2015).

[12] H. Al-Bataineh, Deep Contextualized Pairwise Semantic Similarity for Arabic Language Questions. (2019).

[13] H.Zhang, A survey on the techniques, applications, and performance of short text semantic similarity. Concurrency and Computation: Practice and Experience, 33(5), e5971. (2021).

[14] P.Sunilkumar, Survey on Semantic Similarity. In International Conference on Advances in Computing, Communication and Control (ICAC3) (pp. 1-8). IEEE. (2019).

[15] E. Ismail . Accuracy evaluation of methods and techniques in Web-based short-answer grading systems: a survey. Knowledge and Information Systems, 58(3), 611-650 (2019).

[16] H., Po-Sen, A study of using syntactic cues in short-text similarity measure." Journal of Internet Technology 20.3 839-850. (2019)

[17] I. Inkpen, Semantic similarity of short texts. Recent Advances in Natural Language Processing V, 309, 227- 236. (2009).

[18] Z.Quan, Document Summarization using Word and Part-of-speech based on Attention Mechanism. In Journal of Physics: Conference Series (Vol. 1168, No. 3, p. 032008). IOP Publishing (2019).

[19] S .Awajan, Using Part of Speech Tagging for Improving Word2vec Model. In 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS) (pp. 1-7). IEEE. (2019).

[20] P. Mage, Calculating the similarity between words and sentences using a lexical database and corpus statistics. (2018).

[21] S.Pawade, Subjective answer grader system based on machine learning. In Soft Computing and Signal Processing (pp. 347-355). (2019).

[22] C. Pan. Measuring distance-based semantic similarity using meronymy and hyponymy relations. Neural Computing and Applications, 32(8), 3521-3534 (2020).

[23] Yu. Wang, Understanding short texts through semantic enrichment and hashing. IEEE Transactions on Knowledge and Data Engineering, 28(2), 566-579. (2015).

[24] S. Pawade Subjective answer grader system based on machine learning. In Soft Computing and Signal Processing (pp. 347-355). , (2019).

[25] H.Qu, Interactive knowledge-enhanced attention network for answer selection. Neural Computing and Applications, (2020).

[26] K. Monsefi, Attention-based Convolutional Neural Network for Answer Selection using BERT. In 2020 8th Iranian Joint Congress on Fuzzy and intelligent Systems (CFIS) (pp. 121-126). IEEE. (2020).

[27] M. Fatemi,. BAS: an answer selection method using BERT language model. 1911.01528. (2019).

[28] S. K. Hamed "A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification," J. Comput. Sci., vol. 12, no. 3, pp. 169–177, (2016).

[29] H.Yang, M. Interactive knowledge-enhanced attention network for answer selection. Neural Computing and Applications, 1-17. (2020).

[30] K. Monsefi,. Attention-based Convolutional Neural Network for Answer Selection using BERT. In 2020 8th Iranian Joint Congress on Fuzzy and intelligent Systems (CFIS) (pp. 121-126). IEEE. (2020).

[31] D. Chang. Bert: Pre-training of deep bidirectional transformers for language understanding. (2018).

[32] S.Parreiras, A literature review on question answering techniques, paradigms and systems. Journal of King Saud University-Computer and Information Sciences, 32(6), 635-646. (2020).