Article:

Alshammeri, M orcid.org/0000-0002-4645-3991, Atwell, E orcid.org/0000-0001-9395-3764 and Alsalka, MA orcid.org/0000-0003-3335-1918 (2021) A Siamese Transformer-based Architecture for Detecting Semantic Similarity in the Quran. The International Journal on Islamic Applications in Computer Science And Technology - IJASAT, 9 (4). ISSN 2289-4012

A Siamese Transformer-based Architecture for Detecting Semantic Similarity in the Quran

Menwa Alshammeri[1,2,a], Eric Atwell[2,b], Mhd Ammar Alsalka[2,c]

[1]Jouf University, Sakaka, Saudi Arabia

[2]University of Leeds, Leeds, LS2 9JT, UK

[a]scmhka@leeds.ac.uk , [b]E.S.Atwell@leeds.ac.uk, [c]M.A.Alsalka@leeds.ac.uk

ABSTRACT

Semantic similarity detection is a crucial task in natural language comprehension and plays an important role in many NLP applications such as information extraction, words sense disambiguation, text summarization, and text clustering. This paper focuses on the semantic similarity in the Quran. We propose a Siamese transformer-based architecture for pairwise semantic similarity detection in the Quran. We exploit Arabic pre-trained contextual representations to derive semantically meaningful verse embeddings. We then finetune the twin transformers networks on a semantic similarity dataset drawn from the Quran. We show that our model improves the Quranic semantic similarity measures and performance over previous studies.

Keywords: The Quran, Semantic Similarity, Transformer, Pre-trained Contextual Representations, Siamese Transformer-Networks, AraBERT, SBERT.

1. Introduction

Measuring Semantic Textual Similarity (STS) is an active research area in natural language processing with a wide range of applications such as question answering, document summarization, paraphrase detection. Several models have been proposed to measure semantic similarity depending on its applications: paraphrase detection (Fernando & Stevenson, 2008), text summarization (Aliguliyev, 2009), question-answering systems (Verspoor & MacKinlay, 2012; Gómez-Adorno et al., 2013; Filice et al., 2017). Most available similarity measures were created and used for English texts. Only a few measures were designed mainly for Arabic literature (for a recent survey on Arabic semantic similarity measures see (Alian & Awajan, 2018)).

The task is challenging due to the Arabic Language's complicated morphology, structure, and ambiguity—Furthermore, the absence of relevant resources and tools, especially when dealing with classical Arabic text (El-Deeb et al., 2018). The Holy Quran is a unique Classical Arabic text that encodes subtle religious meanings unrevealed by direct and simple analysis (Alqahtani & Atwell, 2014). This feature of the holy Quran made it the correct text to study semantic similarity between its passages.

A variety of models have been proposed to address the semantic similarity task (Zhao et al., 2014; Wu et al., 2017; Feng et al., 2017; Wang et al., 2017; Tan et al., 2018). Neural networks are favoured over conventional machine learning models because they consistently outperform traditional machine learning models. They also don't rely on linguistic features. Therefore, they can be utilized evenly in languages other than English. The transformer architecture, in particular, has seen significant breakthroughs in deep learning over the last few years. Recent pre-trained contextualized representations, such as ELMo (Peters et al., 2018) and A Bidirectional Encoder Representations from Transformers BERT (Devlin et al., 2019), have significantly increased performance across various NLP tasks, including semantic similarity detection.

The architecture employed in this paper is a special class of neural networks called Siamese architecture; two or more identical networks. The networks are identical because they have the same configuration with the same parameters and weights. Siamese networks have recently been used with success in sentence similarity (Mueller & Thyagarajan, 2016; Neculoiu et al., 2016; Ranasinghe et al., 2019). Recently, Reimers & Gurevych (2019) presented Sentence-Bert (SBERT), a modification of the pre-trained BERT network. SBERT uses Siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity with less computational overhead.

This work leverages both the pre-trained contextualized representations for the Arabic language and the Siamese architecture to derive semantically meaningful verse embeddings and achieve remarkable results in pairwise semantic similarity detection in the Quran. We, indeed, adopt a Siamese sentence-transformer networks structure to yield useful sentence embeddings and form a highly structured space with deep semantic relationships reflected in its geometry. We use a pre-trained BERT model that supports the Arabic Language as a sentence transformer and fine-tunes it on a semantic similarity dataset drawn from the Quran. We experiment with AraBERT (Antoun et al., 2020) as feature extractor models. To fine-tune the model, we create a Siamese network to update the weights so that the derived sentence/ verse embeddings are semantically meaningful and can be compared with cosine similarity.

The paper is organized as follows: Section 2 provides a review on semantic similarity approaches for the Arabic language and the Quran, and a background on Transformers and Pretrained language models and their applications. Section 3 contains information about datasets used. Section 4 presents the proposed Model. Section 5 describe the experimental setup used to train our models. Section 6 presents our results. The architecture is evaluated in Section 7. The paper finishes with conclusions and future directions.

2. Related Work

2.1 Semantic Similarity for the Arabic language and the Quran

Arabic is considered to be low-resourced language, has many dialects, and rich in morphology. Therefore, identifying semantic similarity in Arabic text is not a trivial task. The Quran, as a significant religious text, uses the classical Arabic to its most potential. Therefore, to quantify semantic similarity between its passages, computational models need to incorporate deep semantic analysis and external domain knowledge.

Researchers have proposed numerous models to determine textual semantic similarity. However, few related works were proposed for Arabic (Alian & Awajan, 2018). Some efforts studied the semantic similarity in Arabic (Mohamed et al., 2015; Mahmoud & Zrigui, 2017; Nagoudi & Schwab, 2017; Al-Bataineh et al., 2019; Al-Theiabat & Al-Sadi, 2020), and in the Quran (El-Deeb et al., 2018; Alshammeri et al., 2021; Alsaleh et al., 2021). However, there is a lack of deep learning studies on the topic of STS in the Quran. Bashir et al. (2021) presented a thorough examination of Qur'anic Arabic NLP approaches, tools, and applications. They also outlined open research challenges and future research possibilities. Their survey can act as a useful reference for researchers and practitioners in the field. NLP in the Qur'an is a growing field of study. However, when compared to Arabic NLP, Qur'anic NLP research is still immature and has a lot of potentials (Bashir et al. 2021). This work aims to boost the semantic similarity measures for the Quran.

2.2 Transformers and Pre-trained Language Models

Vaswani et al. unveiled the innovative Transformer model in 2017, which marked a turning point in the field of NLP. Since then, the NLP community has contributed a number of extremely powerful components that can be freely downloaded and used in different models and pipelines. The release of BERT is one of the most recent milestones in this development. Delvin et al. (2019) created BERT (Bidirectional Encoder Representations from Transformers) a pre-trained language model, that set state-of-the-art records for various NLP tasks, including the Semantic Textual Similarity (STS). As a result, the pre-trained BERT model may be fine-tuned with just one additional output layer to produce cutting-edge models for a variety of tasks.

BERT is based on a number of recent innovations in the NLP community, including ELMo (Peters et al., 2018) and the OpenAI transformer (Radford et al., 2018). Unlike Word2vec and Glove, ELMo generates a word embedding based on the context it's used in – to both capture the word meaning in that context as well as other contextual information. ELMo uses a bi-directional LSTM trained on a specific task to be able to create those embeddings. In terms of NLP, ELMo was a significant step toward pre-training. These pre-trained language models are effective tools, and the paradigm has now been applied to other languages.

2.3 SBERT

Despite the huge successes of Transformer-based language models, BERT's design makes it unsuitable for unsupervised tasks such as clustering and semantic similarity search. BRRT uses a cross-encoder; it represents a single sentence or a pair of sentences in one token sequence. It is disadvantageous that no independent sentence embeddings are computed which makes it difficult to derive sentence embedding from BERT. Researchers from the UKP Lab released S-BERT (Sentence-BERT), which modifies the pre-trained BERT network to use Siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be computed using cosine similarity (Reimers & Gurevych, 2019) (see Figure 1). This inspired us to use the Siamese architecture to detect semantic similarity in the Quran. This paper, therefore, proposes a Siamese transformer networks architecture for pairwise semantic similarity detection in the Quran. We use AraBERT transformers that make up the base, and a Siamese set-up is used to fine-tune the models.

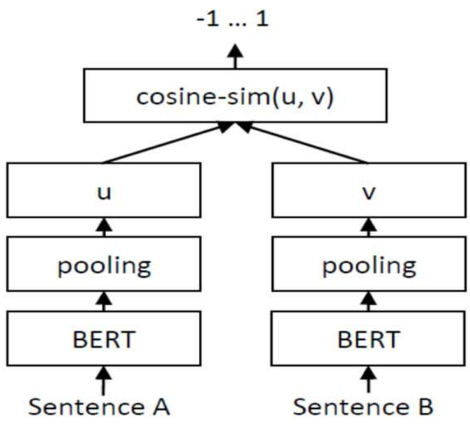

Figure 1: SBERT Siamese network architecture, with regression objective function, for fine-tuning on a STS dataset (Reimers & Gurevych, 2019)

3. Dataset Description

We use a popular benchmark resource that provides pairs of similar verses from the Quran. QurSim (Sharaf & Atwell, 2012) is considered a valuable resource of related pairs of the Holy Quran in which semantically related pairs of verses are linked together. It is regarded as a gold standard resource in analysing relatedness in short texts. Qursim contains 7679 pairs that are related with a degree of relevance 0, 1, or 2. We create a dataset that contains pairs of verses from the Quran with binary labels for semantic similarity/ relatedness: (i) We import the related pairs from Qursim; we tried different combination of pairs given their degree of relevance when training our model (ii)We map the verses' pairs to their text using the Tanzil[1] project. (iii) The dataset undergoes some cleaning to eliminate a total of 372 duplicate records. Examples of such pairs are shown in Table 1. (iiii) We create non-related pairs; randomly generated to be not in Qursim, and (v) We split the data into training and test sets.

Table 1: Examples of duplicated records from Qursim; we keep one of the duplicated pairs such in < 1:5,73:9>, <37:9, 1:5>, and we removed pairs like <30:4, 4: 30> where the verse is related to itself.

No.	Verse1	No.	Verse2
30: 4	فِي بِضْعِ سِنِينَ ۗ لِلَّهِ الْأَمْرُ مِن قَبْلُ وَمِن بَعْدُ ۚ وَيَوْمَئِذٍ يَفْرَحُ الْمُؤْمِنُونَ	30: 4	فِي بِضْعِ سِنِينَ ۗ لِلَّهِ الْأَمْرُ مِن قَبْلُ وَمِن بَعْدُ ۚ وَيَوْمَئِذٍ يَفْرَحُ الْمُؤْمِنُونَ
1: 5	إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ	73: 9	رَّبُّ الْمَشْرِقِ وَالْمَغْرِبِ لَا إِلَٰهَ إِلَّا هُوَ فَاتَّخِذْهُ وَكِيلًا
73: 9	رَّبُّ الْمَشْرِقِ وَالْمَغْرِبِ لَا إِلَٰهَ إِلَّا هُوَ فَاتَّخِذْهُ وَكِيلًا	1: 5	إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ

An Example for the dataset is provided in Table 2 . Finally, we construct the task as predicting the semantic similarity between two verses in a binary classification task.

Table 2: Examples of related pairs from Qursim

Location (Chapter: Verse)	Verse1	Verse2	Relevance
3:142, 29: 2	أَمْ حَسِبْتُمْ أَن تَدْخُلُوا الْجَنَّةَ وَلَمَّا يَعْلَمِ اللَّهُ الَّذِينَ جَاهَدُوا مِنكُمْ وَيَعْلَمَ الصَّابِرِينَ Or do you think that you will enter Paradise while Allah has not yet made evident those of you who fight in His cause and made evident those who are steadfast?	أَحَسِبَ النَّاسُ أَن يُتْرَكُوا أَن يَقُولُوا آمَنَّا وَهُمْ لَا يُفْتَنُونَ Do the people think that they will be left to say, "We believe" and they will not be tried?	2
19: 63, 23: 11	تِلْكَ الْجَنَّةُ الَّتِي نُورِثُ مِنْ عِبَادِنَا مَن كَانَ تَقِيًّا That is Paradise, which We give as inheritance to those of Our servants who were fearing of Allah.	الَّذِينَ يَرِثُونَ الْفِرْدَوْسَ هُمْ فِيهَا خَالِدُونَ Who will inherit Al-Firdaus. They will abide therein eternally.	2

4. Architecture

This paper presents a Siamese transformer-based networks architecture to assess the semantic similarity between Quran verses. Two Siamese transformer networks each process one of the verses in a given pair. Our model incorporates Arabic pre-trained contextualized representation to derive semantically meaningful sentence embeddings and achieve state-of-the-art binary semantic similarity classification results. Furthermore, our model benefits from the Siamese

[1] https://tanzil.net/docs/

network architecture, like in SBERT; two transformer networks have tied weights, to fine-tune the pre-trained model with less computational burden characterizing sentence-pair regression tasks. We experiment with a transformers-based model that were pre-trained for the Arabic Language. The model is AraBERT, where Antoun et al. (2020) pre-trained BERT transformer model (Devlin et al., 2019) for the Arabic Language. So, we apply the sentence pairs classification task on Arabic Quran verses by fine-tuning the non-segmented AraBERT model. Our architecture is depicted in Figure 2.

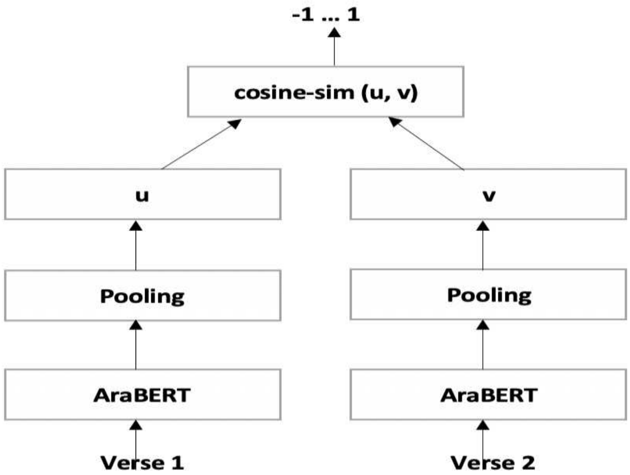

Figure 2: A Siamese transformer networks architecture, with regression objective function, for fine-tuning on Quranic semantic similarity dataset

Each transformer network is composed of a transformer layer and a pooling layer. We define 'bert-base-arabertv02' as the embedding layer that produces contextualized word embeddings for all input tokens in our text. We then define a pooling layer to derive a fixed-sized output representation; we use mean-pooling. Next, we average all contextualized word embeddings produced at the transformer layer to generate the sentence embeddings. Finally, we fine-tune the pre-trained model using our training dataset. When fine-tuning on a semantic similarity dataset, a regression objective function is used, where the cosine similarity between two sentence embeddings, Verse1 and Verse2, is calculated, and mean-squared error loss is used as the loss function.

5. Experiment

We fine-tune sentence-level AraBERT model on an STS dataset. An objective regression function with mean-squared error loss is used. We use the sentence-transformers Python library from the UKP Lab[2]. We use version two of a non-segmented AraBERT model "bert-base-arabertv02". We train our model using our semantic similarity dataset consisting of verses pairs labeled as 1 or 0 for the binary semantic similarity between the two verses.

We run our experiment using a Google Colab notebook to benefit from the free-tier GPU instance to speed up the training. First, we load the training dataset. Then we split the train set for validation data during training while keeping the test set to evaluate the final model. Finally, we experiment with the transformer and set the model’s name to bert-base-arabertv02". The model is built using three modules: a word embedding layer, a mean-pooling layer and a dense layer. We fine-tune our model with an objective regression function for eight epochs. We used a batch size of 32, Adam optimizer with learning rate 2e−5, 10% of the training data is used

[2] https://github.com/UKPLab/sentence-transformers

for warm-up, and evaluation is set to happen every 1000 steps. Our default pooling strategy is MEAN. The last step is to evaluate the model on the STS test set.

6. Results

The results in Table 3 below are for cosine similarity, Manhattan distance, Euclidean distance, and dot-product similarity, as measured by the Pearson correlation and Spearman correlation metrics. According to (Reimer et al., 2016), unlike Pearson correlation, the Spearman correlation metric is more suited for evaluating STS tasks. We, therefore, consider the spearman correlation as representative of the ability to accurately determine whether two verses are similar. Our model achieves a score of 84.96% with AraBERT using the pairs with relevance degree 1 and 2 from Qursim corpus as the similar pairs. Table 4 shows the spearman correlation for cosine similarity using the different combinations of relevance degree for similar pairs for our model.

Table 3: Evaluation of holdout test data on a Quranic Semantic Similarity dataset

Cosine-Similarity:	**Pearson: 0.9163**	**Spearman: 0.8496**
Manhattan-Distance:	**Pearson: 0.8999**	**Spearman: 0.8477**
Euclidean-Distance:	**Pearson: 0.8955**	**Spearman: 0.8471**
Dot-Product-Similarity:	**Pearson: 0.9097**	**Spearman: 0.8495**

Table 4: Spearman correlation scores using different settings of our model and different versions of data

Model	# Epochs	Relevance degree for similar pairs	Spearman correlation Cosine-Similarity
Siamese AraBERT	**4**	**2**	**79.01%**
		1 + 2	**82.32%**
		0 + 1 + 2	**80.71%**
	8	**2**	**78.32%**
		1 + 2	**84.96%**
		0 + 1 + 2	**83.35%**

The results confirm that a Siamese AraBERT networks architecture has the capability to accurately evaluate whether two verses are similar which is represented by the Spearman correlation. Indeed, AraBERT shows exceptional performance on detecting semantic similarity in the Quran.

7. Evaluation

We evaluate our model for predicting similarity using the Quranic semantic similarity dataset. The evaluation metric we use for this task is accuracy and F1-Score. We compute the verses embeddings for each pair using the Siamese AraBERT networks architecture to be used to calculate the cosine distance and determine how semantically similar they are. Using a threshold of 0.60 for the cosine similarity, we consider the pairs with similarity equal or above the threshold to be related (1), otherwise non-related (0). We then compared the actual results (the annotation in the dataset) with the predicated ones (using our model). Our model scored 95% accuracy, and F1-score of 95%. Table 5 shows the confusion matrix and the classification report.

Table 5: Classification Report and confusion matrix

CLASSIFICATION REPORT	**PRECISION**	**RECALL**	**F1 SCORE**	**SUPPORT**	**CONFUSION MATRIX**	**ACTUALLY POSITIVE (1)**	**ACTUALLY NEGATIVE (1)**
0	0.92	0.98	0.95	6236	**Predicted Positive (1)**	**TP 6114**	**FP 122**
1	0.98	0.92	**0.95**	6436	**Predicted**	**FN**	**TN**

					Negative (0)	526	5910
ACCURACY			0.95	12672			

We further report performance of our model on the same dataset, compared against systems of earlier work based on accuracy and F1-score as shown in Table 6.

Table 6: Performance of the Siamese transformer-based networks architecture (AraBERT) compared to Doc2 model

Metric	Siamese AraBERT	AraBERT	Doc2vec
Accuracy	95%	92.1%	76%
F1-Score	95%	85%	54%

7.1 Qualitative Evaluation

We also qualitatively look at examples where our model made correct and wrong predictions. We picked pairs that we have already known are related with some degree of relevance; 1 and 2 from Qursim. We picked non-related pairs as well; we labelled non-related pairs with the relevance of -1. We compared the actual target values with those predicted by our model to identify what type of predication is made. The results are reported in Table 7. Studying the results provides insights on where our model failed in prediction. It also suggests another area of improvement by using different data, as we generated non-related pairs randomly.

Table 7: Performance of the Siamese AraBERT architecture on the Quranic semantic similarity dataset

Location (Chapter: Verse)	Verse 1	Verse2	Actual Value	Predicted value (%)	Prediction[3]
(8:39, 2: 193)	وقاتلوهم حتى لا تكون فتنة ويكون الدين كله لله فإن انتهوا فإن الله بما يعملون بصير And fight them until there is no fitnah and [until] the religion, all of it, is for Allah. And if they cease - then indeed, Allah is Seeing of what they do.	وقاتلوهم حتى لا تكون فتنة ويكون الدين لله فإن انتهوا فلا عدوان إلا على الظالمين Fight them until there is no [more] fitnah and [until] worship is [acknowledged to be] for Allah. But if they cease, then there is to be no aggression except against the oppressors.	2	99.51	TP
(6: 142, 35: 6)	ومن الأنعام حمولة وفرشا كلوا مما رزقكم الله ولا تتبعوا خطوات الشيطان إنه لكم عدو مبين	إن الشيطان لكم عدو فاتخذوه عدوا إنما يدعو حزبه ليكونوا من أصحاب السعير	1	56.44	FN

[3] The Prediction can be TP (true positive), FP (false positive), TN (true negative), or FN (false negative). We define positive and negative to be similar and non-similar respectively. True means the actual value matches the predicted value, and negative means the predicted and actual values do not match. The actual value represents the actual similarity from the semantic similarity dataset (it could be 2 or 1 if similar, and -1 if non-similar), while the predicted value represents the similarity score computed based on the proposed model. For example, TP means the predication was positive (similar with score >= 60) and the actual values was also positive (relevance is 2 or 1). FP means the predication was positive (similar) and the actual value was negative (non-similar).

	And of the grazing livestock are carriers [of burdens] and those [too] small. Eat of what Allah has provided for you and do not follow the footsteps of Satan. Indeed, he is to you a clear enemy.	Indeed, Satan is an enemy to you; so, take him as an enemy. He only invites his party to be among the companions of the Blaze.			
(55:41,42: 39)	يعرف المجرمون بسيماهم فيؤخذ بالنواصي والأقدام	والذين إذا أصابهم البغي هم ينتصرون	-1	21.35	TN
	The criminals will be known by their marks, and they will be seized by the forelocks and the feet.	And those who, when tyranny strikes them, they defend themselves.			
(33:24, 47:31)	ليجزي الله الصادقين بصدقهم ويعذب المنافقين إن شاء أو يتوب عليهم إن الله كان غفورا رحيما	ولنبلونكم حتى نعلم المجاهدين منكم والصابرين ونبلو أخباركم	-1	60.00	FP
	That Allah may reward the truthful for their truth and punish the hypocrites if He wills or accept their repentance. Indeed, Allah is ever Forgiving and Merciful.	And We will surely test you until We make evident those who strive among you [for the cause of Allah] and the patient, and We will test your affairs.			

8. Conclusion

This paper presented a verse-embedding using Siamese AraBERT-networks for fast and efficient semantic similarity detection in the Quran. We proposed a Siamese transformer-based architecture where two networks have tied weights. Our architecture starts with pre-trained AraBERT model to achieve high performance, and a Siamese set-up is used to fine-tune the models on a semantic similarity dataset drawn from the Quran. Our model achieved an 84.96% Spearman correlation representing its ability to assess whether two verses are similar. Furthermore, our model sets a high-performance record of 95% F1-score on the Quranic semantic similarity dataset.

Many research potentials are using transformer-based sentence embedding across a wide range of NLP tasks and leveraging the paradigm of pre-trained language models and the application of transformers to the Arabic language and the Quran. The only limitation here is the availability of labelled datasets for fine-tuning. One potential is to create datasets drawn from the Quran and relevant knowledge resources to support training and fine-tuning transformers.

This work can be an excellent starting point for many potential types of research on modelling the semantic similarity/ relatedness in the Quran and extracting the embedded knowledge. For example, we may use our model for topic identification and unsupervised tasks such as clustering based on the semantic similarity of sentence/verse embeddings, providing a deep learning alternative to traditional models such as Latent Dirichlet Allocation. We would experiment using tBERT, a topic-informed BERT-based architecture for pairwise semantic similarity detection, using AraBERT as the base.

Acknowledgements (Optional)

The first author is supported by a PhD scholarship from the Ministry of Higher Education, Saudi Arabia. The author is grateful for the support from Jouf University for sponsoring her research.

References

Akila, M. D., & Jayakumar, C. (2014). Semantic Similarity-A Review of Approaches and Metrics. International Journal of Applied Engineering Research, 9(24), 27581-27600.

Alian M. & Awajan A., (2018). Arabic Semantic Similarity Approaches – Review. International Arab Conference on Information Technology (ACIT)(pp. 1-6). doi: 10.1109/ACIT.2018.8672665.

Aliguliyev, R. M. (2009). A new sentence similarity measure and sentence based extractive technique for automatic text summarization. Expert Systems with Applications, 36(4), 7764-7772.

Antoun, W., Baly, F., & Hajj, H. (2020). Arabert: Transformer-based model for arabic language understanding. arXiv preprint arXiv:2003.00104.

Alqahtani, M., & Atwell, E. (2016, June). Arabic Quranic search tool based on ontology. In International Conference on Applications of Natural Language to Information Systems (pp. 478-485). Springer, Cham.

Alsaleh, A. N., Atwell, E., & Altahhan, A. (2021, April). Quranic Verses Semantic Relatedness Using AraBERT. In Proceedings of the Sixth Arabic Natural Language Processing Workshop (pp. 185-190). Leeds.

Alshammeri, M., Atwell, E., & ammar Alsalka, M. (2021). Detecting Semantic-based Similarity Between Verses of The Quran with Doc2vec. Procedia Computer Science, 189, 351-358.

Al-Bataineh, H., Farhan, W., Mustafa, A., Seelawi, H., & Al-Natsheh, H. T. (2019, November). Deep contextualized pairwise semantic similarity for arabic language questions. In 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 1586-1591). IEEE.

Al-Theiabat, H., & Al-Sadi, A. (2020). The inception team at NSURL-2019 task 8: Semantic question similarity in Arabic. arXiv preprint arXiv:2004.11964.

Bashir, M. H., M Azmi, A., Nawaz, H., Zaghouani, W., Diab, M., Al-Fuqaha, A., & Qadir, J. (2021). Arabic Natural Language Processing for Qur'anic Research: A Systematic Review.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

El-Deeb, R., Al-Zoghby, A. M., & Elmougy, S. (2018). Multi-corpus-Based Model for Measuring the Semantic Relatedness in Short Texts (SRST). Arabian Journal for Science & Engineering (Springer Science & Business Media BV), 43(12).

Feng, W., Wu, Y., Wu, W., Li, Z., & Zhou, M. (2017, August). Beihang-msra at semeval-2017 task 3: A ranking system with neural matching features for community question answering. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (pp. 280-286).

Fernando, S., & Stevenson, M. (2008, March). A semantic similarity approach to paraphrase detection. In Proceedings of the 11th annual research colloquium of the UK special interest group for computational linguistics (pp. 45-52).

Filice, S., Da San Martino, G., & Moschitti, A. (2017, August). Kelp at semeval-2017 task 3: Learning pairwise patterns in community question answering. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (pp. 326-333).

Gómez-Adorno, H., Pinto, D., & Vilarino, D. (2013, June). A question answering system for reading comprehension tests. In Mexican Conference on Pattern Recognition (pp. 354-363). Springer, Berlin, Heidelberg.

Mahmoud, A., & Zrigui, M. (2017, November). Semantic similarity analysis for paraphrase identification in Arabic texts. In Proceedings of the 31st Pacific Asia conference on language, information and computation (pp. 274-281).

Mohamed, R., Ragab, M., Abdelnasser, H., El-Makky, N. M., & Torki, M. (2015, June). Al-Bayan: A knowledge-based system for Arabic answer selection. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) (pp. 226-230).

Mueller, J., & Thyagarajan, A. (2016, March). Siamese recurrent architectures for learning sentence similarity. In Proceedings of the AAAI conference on artificial intelligence(Vol. 30, No. 1).

Nagoudi E. B., & Schwab D. (2017, April). Semantic Similarity of Arabic Sentences with Word Em- beddings. In proceedings of the Third Arabic Natural Language Processing Workshop, Apr (pp.18 – 24). hal-01683485.

Neculoiu, P., Versteegh, M., & Rotaru, M. (2016, August). Learning text similarity with siamese recurrent networks. In Proceedings of the 1st Workshop on Representation Learning for NLP (pp. 148-157).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding with unsupervised learning. OpenAI.

Ranasinghe, T., Orăsan, C., & Mitkov, R. (2019, September). Semantic textual similarity with siamese neural networks. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019) (pp. 1004-1011).

Reimers, N., Beyer, P., & Gurevych, I. (2016, December). Task-oriented intrinsic evaluation of semantic textual similarity. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers(pp. 87-96).

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

Sharaf, A. B. M., & Atwell, E. (2012, May). QurSim: A corpus for evaluation of relatedness in short texts. In LREC (pp. 2295-2302).

Tan, C., Wei, F., Wang, W., Lv, W., & Zhou, M. (2018, January). Multiway Attention Networks for Modeling Sentence Pairs. In IJCAI (pp. 4411-4417).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems(pp. 5998-6008).

Verspoor, K., & MacKinlay, A. (2012). Simple similarity-based question answering strategies for biomedical text.

Wang, Z., Hamza, W., & Florian, R. (2017). Bilateral multi-perspective matching for natural language sentences. arXiv preprint arXiv:1702.03814.

Wu, G., Sheng, Y., Lan, M., & Wu, Y. (2017, August). ECNU at semeval-2017 task 3: Using traditional and deep learning methods to address community question answering task. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (pp. 365-369).

Zhao, J., Zhu, T., & Lan, M. (2014, August). Ecnu: One stone two birds: Ensemble of heterogenous measures for semantic relatedness and textual entailment. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 271-277).