International Journal on Islamic Applications in Computer Science And Technology, Vol. 11, Issue 1, March 2023, 09- 18

Islamic Question Answering Systems Survey and Evaluation Criteria

Sarah Alnefaie[1], Eric Atwell[2], and Mohammad Ammar Alsalka[3]

[1,2,3] University of Leeds, Leeds, UK
[2] King Abdulaziz University, Jeddah, Saudi Arabia
[1]scsaln@leeds.ac.uk, [2]e.s.atwell@leeds.ac.uk, [3]m.a.alsalka@leeds.ac.uk

Abstract
Many researchers built Islamic question-answering systems which find the answer to the question from the Quran, Hadith, or Fatwa text using only efficient retrieval techniques. However, it is challenging to answer all kinds of questions due to the current shortcomings in natural language processing tools. Therefore, in this paper, we review the Islamic question-answering systems that can answer all kinds of questions by building a questions and answers corpus and then using the retrieval technique or pre-training model to answer the user's question. After that, we use thirteen evaluation criteria, such as the search approaches and the system scope, to evaluate these systems. We can conclude from this survey that there are flaws in the existing systems, such as all these systems being unavailable and can answer a limited number of questions.

***Keywords*:** Islamic Search Tool, Islamic Information Retrieval, Quranic NLP.

1. Introduction

When Muslims have a question, they consult Islamic experts to answer it with Quran or Hadith evidence. They usually use the search engine to find the answer, but with the massive increase in internet information, the search engine retrieves many sites that may contain the answer. The user usually needs time and effort to read and verify all these answers. Therefore, there is a need for a Question Answering (QA) system that can extract reliable and fast answers to the user's questions.

Many systems were built to retrieve knowledge from holy books. These systems can be divided into two types. The first type focused on retrieving data directly from the Quran or Hadith texts can be called systems based on retrieval techniques?. They relied entirely on the power of NLP techniques to find the answer. In contrast, the knowledge base systems focus on building a dataset containing the most significant number and different types of questions and their answers from Islamic resources such as the Quran and Hadith and then using retrieval techniques or a pre-trained deep learning model to extract the correct answers.
One of the most significant weaknesses in the existing systems based on retrieval techniques is answering only the factoid questions (Maraoui et al., 2021). Therefore, this paper will focus on the knowledge base systems that answer all kinds of possible questions. These systems build a Questions and Answers (Q&A) corpus and then use traditional retrieval techniques or a pre-trained deep learning model to retrieve the answers. This paper is structured as follows: In the second section, we will survey the existing knowledge base systems. The third section will present the methodology for evaluating these systems. The last section will discuss the conclusion.

2. Literature Review

Many Islamic knowledge base systems have been built, which can divide into systems that use traditional retrieval techniques and a pre-trained deep learning model:

2.1 Knowledge Base System using Traditional Retrieval Technique

Saeedi et al. (2014) introduced a Persian QA system called Quranjooy for Quran. They gathered 6000 Q&A from credible websites to develop a knowledge base. This system has three stages. First, the system applied the question analysis stage. Second, four modules were applied parallelly to retrieve the answer from the knowledge base. Finally, the outputs of these modules merged to get the correct answer. The modules in the second stage are Named-entity recognition (NER), verse finder, tabular, and ontology-based. Each module considered the most relevant answer as an output. The NER module can answer nine types of questions. For example, if the question type is a date, this module will focus the search process on words tagged as date using the NER tool. The second model can answer the question about the verse address and the frequency of a specific term in the Quran. Questions about interpretation are answered using the tabular module. The last module's idea was building an ontology to retrieve the answer. In addition, Heidaria et al. (2014) implemented this system using the GATE framework. However, this system has many flaws. It faces problems with Persian pre-processing tools. In addition, the coverage of the ontology is limited. The system's ability to answer is limited to a certain number of questions in some question types. This system is for Persian questions.

In addition, Sheker et al. (2016) developed two Fatwas QA systems, an ontology-based system and a synonym-set system. Usually, Muslims ask Islamic experts questions about their opinion on many daily issues that face them, which is called Fatwa. The two knowledge resources were created for these systems. First, a corpus of 1094 Q&A in the prayers Fatwas area was constructed based on Ibn Baz's book (Ibn Baz, Abdul Aziz and AlUthaymeen, Muhammad and Al-Madkhalee, 2003). Second, an ontology has been built based on this corpus.

The architecture of these systems involves three stages: question pre-process, question analysis, and question expansion. The encoding and normalization operation was implemented to pre-process the question. The question analysis phase aims to return the corpus questions similar to the user question. First, the similarity measures using Cosine and Jaccard produces a two list of the candidate questions. Afterwards, the candidate questions rank based on the frequency among the lists. Finally, the most relevant question is retrieved from the dataset, and its answer is displayed to the user. The architecture of the two systems is the same, but there is a difference only in the phase of question expansion. The question expansion reformulates the query by adding a synonym. The first system uses the synonyms found in the new ontology, while the second uses WordNet synonyms. Thirty questions were used to evaluate the systems proposed by experts. The performance of an ontology-based system is better compared with the synonym-set system. The first system achieves 91% F-measure, 90% recall, and 92% precision, while the second obtain 65%, 59%, and 72%, respectively.

However, there are limited concepts in the ontology and weaknesses in the Arabic wordnet, which affect the performance. Additionally, the system is unavailable and can answer only a limited number of questions about prayer. The answer is a Fatwa with expert words without verse or Hadith evidence.

Furthermore, Hamoud and Atwell (2016a, 2016b) proposed answering the Arabic and English Quran questions using the knowledge base. This system consists of a twocomponents: a corpus

and a QA model. The 1500 Q&A pairs were collected from different resources to build a corpus. To enhance the data quality, they cleaned the data. Next, they formatted the corpus in the comma-separated value (CSV) format. After that, the system's performance improved by applying data redundancy techniques, such as paraphrasing the questions in different ways and contexts. This system accepts the question from the user. Next, understand the question by implementing the pre-processing operation. Then, they matched the question with the dataset using a text-based approach. Afterwards, the matched questions were scored and ranked to select the most matched question using a similarity measure. Lastly, show the answer to the users. They used Python natural language toolkit (NLTK). To evaluate this system, they used 63 English and 71 Arabic questions from the Muslims in the university. The Arabic version of this system achieved 79% in precision and 76% in recall, while the English obtained 75% and 73% in precision and recall, respectively.

However, this system is unavailable and uses the text-based retrieval approach, which has many limitations. It answers only Quran questions. This system's performance is affected by the corpus's quality and quantity of data. Unfortunately, there are many flaws in this corpus. Firstly, many questions in this corpus were unreliable because a scholar had not checked them. In addition, they did not mention the reference for each answer. Additionally, many answers do not have evidence from the Quran. Furthermore, the data size is small.

Moreover, Adany (2017) recommended a QA system with two stages for the Quran domain. The first stage is constructing a corpus for Al-Baqarah and Al-Fatiha chapters only, while the second is designing a prototype for QA systems based on this corpus. They created a corpus of 263 Quran questions divided into two categories: a gold standard evaluation dataset and a knowledge base for this study.

There are four steps in this prototype, which was implemented by java. First, it accepts the user’s question and pre-processes it by removing the diacritics and the stopping word. Then, it matches the user’s question with the Q&A corpus. If it exists, then the answer will retrieve and display to the user. If the question does not exist in the corpus, then a search process will apply to find the answer from the Quran text. Finally, If the answer is not found, the question will be asked later to an expert and added to the dataset.

Nevertheless, this system suffers from many shortcomings, such as the system can answer only 263 questions about sura Al-Fatihah and sura Al-Baqarah. Additionally, this system is unavailable for public use and fails to answer questions which are similar to the system's questions in meaning but in different terms. The expert did not comprehensively verify the answers in this system.

2.2 Knowledge Base System using Pre-Trained Model

This section will review the recent Quran transformer-based QA models containing two phases: training and testing. First, these models train on the training set of the Q&A dataset. Next, in the testing phase, these systems take from the test-set two inputs: a Quranic passage and a Modern Standard Arabic (MSA) question, and then retrieve the top answers from the list, which ranked from the highest probability of being correct to the least. The answer is part of the passage. We will show these models in the following:

ElKomy and Sarhan (2022) proposed a Quran QA system using the Qur'anic Reading Comprehension Dataset (QRCD) [1] that consists of a total of 1373 question-passage-answer triplets (Malhas et al., 2022). They employed the training and development sets of QRCD to train five Arabic BERT models: MARBERT, QARiBBase, AraBERTv02Large, ARBERT, and AraBERTv02Base. Then, the models were implemented using the test set of the QRCD. After that, they applied an ensemble approach to select the top fifteen answers using the majority votes of these models. Finally, they improved the answers' quality by implementing the post-processing method. This method consists of three steps: First, dropping or extending tokens to solve the problem of broken words. Second, applying an algorithm eliminates redundancy between answers. Third, eliminating the uninformative answers from the list, such as if the entire answers are stop words. Only the top five possible answers in the new list will be considered. They get a Partial Reciprocal Rank (pRR) of 56.7%, Exact Match (EM) of 26.9%, and F1@1 of 50.2%.

Additionally, Ahmed et al. (2022) recommend using two pre-trained Arabic QA models: AraElectra-Artydiqa and AraElectra-ARCD. First, they tried to enhance the performance by augmenting the QRCD dataset with 657 Q&A using several manual methods, including reformulating the questions by changing the term order or adding synonyms. Then, the system followed the following steps: First, the data was pre-processed by removing stopwords and punctuation. After that, they fine-tuned the two models using these datasets. Next, each model was run using the test part of QRCD. Finally, the hybrid method was applied by selecting the high-score answer from the two models. The system will return the best five answers for each question. The AraElectra-Artydiqa outperforms the AraElectra-ARCD and the hybrid model with: 0.559 pRR, 0.513 F1@1, and 0.244 EM.

Furthermore, Mostafa and Mohamed (2022) developed a QA system based on the idea that the deep learning model must fine-tune using a large corpus to get good results. Their method consists of many steps: First, the MARBERTv2 and ARAELECTRA Models are fine-tuned on the training set of QRCD. Then, they applied the two models to the development set of QRCD. Next, the model with the highest score was selected. Afterwards, they used a different dataset: Ar-TyDi QA, and the merged dataset of Arabic-SQuAD and ARCD, to fine-tune the chosen model and save the model's weight. After that, they used the previous model's weight to fine-tune the chosen model on the training set of the QRCD. Subsequently, the test dataset of QRCD was employed to apply the chosen model. Finally, they address the data imbalance issue by using different loss functions. The ARAELECTRA significantly outperformed MARBERTv2. Therefore, they use the ARAELECTRA model to implement the rest steps. The system achieved 54.6% in terms of pRR.

Wasfey et al. (2022) suggested different methods to build the system. In the first approach, they fine-tuned the AraBERT model with QRCD. Secondly, the Qur'anBERT was created by fine-tuning the AraBERT using Quran verses. Then, the Qur'anBERT fine-tuned using Q&A datasets such as QRCD, Arabic Squad, and Arabic Question Answers from the Holy Qur'an dataset (AQAQ), which the authors created. They ran the models with the QRCD as the second step in both approaches. While in the third method, they applied a voting approach between different models, such as Arabertv02 and AraElectra to select the best answer for the QRCD After they trained them by QRCD and AQAQ. The last approach was the best, with 0.528 pRR, 0.256 EM, and 0.507 F1@1.

[1] https://gitlab.com/bigirqu/quranqa

In addition, Premasiri et al. (2022) conducted three experiments to build a QA system. First, they used the QRCD to fine-tune and run seven individual pre-trained models: AraELECTRA-generator, AraBERTv2, AraELECTRA-discriminator, camelbert-ca, camelbert-mix, mbert-uncased and mbert-cased. The order of three best-performing models was camelbert-mix, camelbert-ca, and AraELECTRA. Secondly, they re-implemented the experiment with the self-ensemble approach on each model by training the same model using five random seeds and then applying an ensemble algorithm to obtain one output. The effect of the self-ensemble approach on the model is very small, as some slightly worsened and some models improved their performance, such as AraELECTRA-discriminator, which is the best among other models. Third, they chose the best three models from the previous experience and then applied the learning transfer approach using the SOQAL dataset. Its idea is to take advantage of rich resources by training the model on it and then memorizing its weight and using that weight when running the model to a small dataset. After that, the second experiment was re-implemented by running AraELECTRA-discriminator but with the saved weight. The result of the AraELECTRA-discriminator model is the highest, with 0.495 pRR, 0.227 EM, and 0.476 F1@1.

Mellah et al. (2022) suggested using the mT5 transformer-based QA system. First, they removed stop words. Then, they fine-tuned three different size models of mT5: Base, Large, and XL on the training set of QRCD and ran them on the development set. After that, the highest-performing model was applied to the test set. The best model was mT5-XL which achieved 0.43 in pRR, 0.197 in EM, and 0.399 in F1@1.

Further, Keleg and Magdy (2022) proposed a QA system based on an Arabic BERT model. First, they fine-tuned and implemented the CAMELBERT-MSA and the CAMELBERT-CA individually using the QRCD dataset. The CAMELBERT-CA model outperformed the CAMELBERT-MSA. After that, they tried to improve the CAMELBERT-CA by stemming the text and embedding the NER tags in the text before applying the CAMELBERT-CA. The best score was the pRR of 0.40, obtained by vanilla CAMELBERT-CA. In addition, Alsaleh et al. (2022) developed a system that answers Arabic Quran text questions using AraBERT, CAMeL-BERT, and ArabicBERT pre-trained language models. These models were trained using the QRCD training dataset and then run using the test set. The experiments showed that the AraBERT V0.2 model achieved better results than the other models, with a 0.418 F1 score, 0.445 pRR score, and 0.16 EM score.

Moreover, Aftab and Malik (2022) proposed developing a QA system using one of the three models: a baseline BERT, Bert-Base-Arabic, and Bert-Base-Multilingual-Uncased. First, they used the QRCD to train the three models. The results showed that Bert-Base-Arabic is the best. After that, they enhanced the system by applying two techniques: regularization and data augmentation. The regularization technique's benefit is preventing overfitting, while the benefit of data augmentation is to enlarge the data. They implement a regularization technique called Decoupled Weight Decay. In data augmentation, they selected a set of only 473 out of 1224 Q&A pairs from AQQAC that follow the format of the QRCD dataset. Their system got 0.308, 0.088, and 0.268 in pRR, EM, and F1@1, respectively. At the same time, the worst model recommended by Singh (2022) that used a system based on the AraBERT achieved poor results at 0.191 pRR, 0.042 EM, and 0.091 F1@1.

The comparison of the best Quran QA pre-trained models is shown in Table 1.

Table 1: Comparing the Quran's pre-trained models.

Paper	Approach
(ElKomy and Sarhan, 2022)	1- Fine-tune the five Arabic BERT models: MARBERT, QARiBBase, AraBERTv02Large, ARBERT, and AraBERTv02Base using the training and development sets of the QRCD. 2- Apply the five models using the test dataset. 3- Implement the majority vote method to select the top fifteen answers. 4- Use the post-process method for the selected answers and return the best five answers.
(Ahmed et al., 2022)	1- Augmented the dataset using manual methods. 2- Pre-process the merged dataset of the augmented data and QRCD by removing stopwords and punctuation. 3- Fine-tune the AraElectra-Artydiqa model using the merged data. 4- Apply the models on the test dataset to extract the best five answers for each question.
(Mostafa and Mohamed, 2022)	1- Fine-tune the ARAELECTRA model using Ar-TyDi QA and the merge of Arabic-SQuAD and ARCD. 2- Fine-tune the model for the training set of QRCD using the previous model's weights. 3- Apply the model to the test dataset of QRCD to retrieve the top five answers. 4- Use different loss functions to address the data imbalance issue.
(Wasfey et al., 2022)	1- Fine-tune the AraBERTv02Base and AraElectra models using the AQAQ and the training and development set of QRCD. 2- Apply the model using the test dataset to return the best five answers. 3- Applied a voting approach between the two models to select the best answers
(Premasiri et al., 2022)	1- Train the AraELECTRA-discriminator on the SOQAL dataset and then save the weight. 2- Train the model on the QRCD dataset using the same weight. 3- Run the model on QRCD using five random seeds. 4- Apply an ensemble algorithm to obtain the best answers.
(Mellah et al., 2022)	1- Remove stop words from the QRCD dataset and re-format the dataset. 2- Fine-tune mT5-XL by QRCD. 3- Apply mT5-XL on the test set and retrieve the answers.
(Keleg and Magdy, 2022)	Fine-tune and implement the CAMELBERT-CA model by QRCD.
((Alsaleh et al., 2022)	1-Fine-tune the AraBERTv02Large on the QRCD. 2- Apply the model to the development set of QRCD using different parameter values and choose the best one. 3- Apply the model with the best parameters values to the test dataset of QRCD.
(Aftab and Malik, 2022)	1- Augment the data using 473 Q&A from AQQAC. 2- Implement the weight-decay technique to train the model in a more general way. 3- Fine-tune the Bert-Base-Arabic model using QRCD and AQQAC. 4- Apply the model to the QRCD test set and find the answers.
(Singh, 2022)	Fine-tune the AraBERT model using the trainig set of QRCD the run it useing the test set.

3. Methodology

The objective of this section is to evaluate the systems mentioned in the previous section to find their flaws and shortcomings. The evaluation process is based on selecting the appropriate criteria and then using them to evaluate these systems.

3.1 The Evaluation Criteria of the Islamic Knowledge Base System

Alqahtani and Atwell (2017) used several criteria for assessing the Quran search systems based on retrieval techniques. Some of these criteria and other criteria are used in this study to assess Islamic knowledge base QA systems as the following:

1. Search Approaches:
   A. Traditional Retrieval Technique:
      I. Text-based approach:
         a. Keyword-matching approach: retrieves the document which involves any query terms.
         b. Morphological-based approach: extracts the document that contains any query terms root.
      II. Semantic approach: returns the documents with the same meaning as the query terms.
         a. Ontology-based approach: matches the question with the ontology concepts, extracts them and adds them to the query.
         b. CLIR approach: first, the question is translated into another language, and then the answers that include the translation word are retrieved.
         c. Synonym-set approach: expands the query with synonyms from external resources and matches the answer with this new query.

B. Pre-Trained Model:

I.MARBERT	II.AraBERT	III.AraElectra-ARCD
IV.MARBERTv2	V.AraBERTv02Base	VI.AraELECTRA-discriminator
VII.QARiBBase	VIII.AraBERTv02Large	IX.Bert-Base-Arabic
X.mT5-XL	XI.ARAELECTRA	XII.CAMELBERT
XIII.ARBERT	XIV.AraElectra-Artydiqa	XV.ArabicBERT

2. The System Language:
   Arabic (A) English (B) Persian (P)

3. Question Analysis Process: analyzing the meaning of the question by:
   Normalization (N) POS (P) Tokenization (T)
   Lemmatization (L) Removing the stop words (RS) Removing the punctuation (RP)
   Encoding (E) Removing the diacritics (RD)

4. Knowledge Base Size: The number of Q&A pairs in this knowledge base.

5. Scope of the System Search:
   Quran (Q) Hadith (H) Tafsirs (T) Fatwa (F)

6. Model Coverage Area: Scope of the
   The whole book (All) Part of the book (Part)

7. System Availability: is this system available to the public?
   Yes (Y) No (N)

8. Corpus Availability: is this corpus available for the users?
   Yes (Y) No (N)

9. Answer Type: the nature of the answer text retrieved from this system
   Plain text. (PT) Verse. (V) Verse information (verse number and sura name). (VI)
   Hadith (H) Part of the verse (PV)

10. Evaluation Metrics:
    F-measure. Precision EM pRR F1@1 Recall

11. Experiment Dataset: How to construct the dataset and its size

12. Experiment Result of this system

13. Programming Language used to implement this model

3.2 Comparing the Existing Islamic Knowledge Base System

This study uses the previous thirteen evaluation criteria to compare the existing knowledge base QA systems about religious texts, as shown in Table 2.

Table 2: Comparing the knowledge base question answering systems

	1. The Search Approaches	2.The System	3.Question Analysis	4.Knowledge Base Size	5.Scope of the System	6.Search Domain	7. System Availability	8.Corpus Availability	9. Answer Type	10.Evaluation Metrics	11.Experiment Dataset	12.Experiment Result	13.Programming Language
(Saeedi et al., 2014b)	A. I. a and A. II. a	P	N, P, T, and L	6000 Q&A pairs	Q	Part	N	N	PT	F-measure, Precision, and Recall	852 questions that two experts created	Precision: 56%	GATE framework
(Sheker et al., 2016b)	System 1: A. I. a and A. II. C System 2: A. I.a and A. II. a	A	N and E	1094 Q&A pairs	F						30 questions that were proposed by Expertise	Precision: 92%, Recall: 90% F-measure: 91%	-
(Adany, 2017) **and** (Hamdelsayed and Atwell, 2016)	A. I. a		RS, RD, and T	263 questions with many answers	Q				V and VI	-	-	-	Java standard edition version 8
(Hamoud, 2017)	A. I. b	A and E	N, T, RS, and RP	1500 Q&A pairs	Q and T				PT, V, VI, and PV	Precision and Recall	71 Arabic questions and 63 English questions were gathered from the Muslim university group	Arabic system Precision: 79% Recall: 76% English system Precision: 75% Recall: 73%	Python NLTK natural language toolkit
(ElKomy and Sarhan, 2022)	B.I B.VII B.XIII B.V B.VIII	A		861 triplets of the QRCD	Q			Y	PV	pRR, EM, and F1@1	348 triplets of the QRCD	pRR: 56.7% EM: 26.9% F1@1:50.2%	
(Ahmed et al., 2022)	B.XIV		RS, PR	657 triplets of the augmented data + 861 triplets of the QRCD								pRR: 55.9% EM: 24.4% F1@1:51.3%	
(Mostafa and Mohamed, 2022)	B.XI			861 triplets of the QRCD								pRR: 54.6% EM: 23.9% F1@1:52.5%	
(Wasfey et al., 2022)	B.V B.XI			732 triplets of the AQAQ + 861 triplets of the QRCD				Part of the dataset Y and other N				pRR: 52.8% EM: 25.6% F1@1:50.7%	
(Premasiri et al., 2022)	B.VI			861 triplets of the QRCD				Y				pRR: 49.5% EM: 22.7% F1@1:47.6%	
(Mellah et al., 2022)	B.X		RS									pRR: 43.0% EM: 19.7% F1@1: 39.9%	
(Keleg and Magdy, 2022)	B.XII											pRR: 40.0% EM: 15.1% F1@1:38.2%	
(Aftab and Malik, 2022)	B.IX			473 triplets of the AQQAC + 861 triplets of the QRCD				Part of the dataset Y and other N				pRR: 30.8% EM: 8.8% F1@1:26.8%	
(Singh, 2022)	B.II			861 triplets of the QRCD				Y				pRR: 0.19% EM: 0.042% F1@1:9.1%	
(Alsaleh et al., 2022)	B.VIII											pRR: 44.5% EM: 16.0% F1@1:41.8%	

4. Conclusion

This paper surveys the existing Islamic knowledge base question-answering systems and assesses them using thirteen metrics. Many shortcomings have been found. First, the majority of the existing studies focus on answering Quran questions. Only one study answered the Fatwa questions. There is no study answering Hadith questions. Second, the corps size for these systems does not exceed two thousand questions and answers, which covers a maximum of 2000 verses out of 6,236 verses. However, there is a relationship between the system performance and the number of answers available. When the number of answers in the system increases, the probability of correctly answering the questions increases. Third, All the systems are unavailable to the public. Fourth, there is only one system in English, and the rest are in Arabic. Fifth, credibility is essential in this system, but some systems do not mention the reference for the answer, or a religious expert did not review their dataset. As a result, there is a need to build a reliable Islamic Q&A system that is available in several languages.

References

Adany M. A. 2017. *An automatic question answering system for the Arabic Quran*. Ph.D. thesis, Sudan University of Science and Technology.

Adany, M.AH. 2017. *An automatic question answering system for the Arabic Quran*. Sudan University of Science and Technology.

Aftab, E. and Malik, M.K. 2022. eRock at Qur'an QA 2022: Contemporary Deep Neural Networks for Qur'an based Reading Comprehension Question Answers *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Ahmed, B.H., Saad, M.K. and Refaee, E.A. 2022. QQATeam at Qurán QA 2022: Fine-Tunning Arabic QA Models for Qurán QA Task *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Alqahtani, M. and Atwell, E. 2017. Evaluation criteria for computational Quran search. *International Journal on Islamic Applications in Computer Science And Technology*. **5**(1), pp.12–22.

Alsaleh, A., Althabiti, S., Alshammari, I., Alnefaie, S., Alowaidi, S., Alsaqer, A., Atwell, E., Altahhan, A. and Alsalka, M.A. 2022. LK2022 at Qur'an QA 2022: Simple Transformers Model for Finding Answers to Questions from Qur'an *In*: *Proceedings of the OSACT 2022 Workshop*., pp.120–125.

ElKomy, M. and Sarhan, A. 2022. TCE∗ at Qur'an QA 2022:Arabic Language Question Answering Over Holy Qur'an Using a Post-Processed Ensemble of BERT-based Models *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Hamdelsayed, M.A. and Atwell, E. 2016. Islamic applications of automatic question-answering.

Hamoud, B. and Atwell, E. 2016a. Quran question and answer corpus for data mining with WEKA *In*: *2016 Conference of Basic Sciences and Engineering Studies (SGCAC)*. IEEE, pp.211–216.

Hamoud, B. and Atwell, E. 2016b. Using an Islamic Question and Answer Knowledge Base to answer questions about the holy Quran. *International Journal on Islamic Applications in Computer Science And Technology*. **4**(4), pp.20–29.

Hamoud, B.I. 2017. *A Question Answering System Design about the Holy Quran*. Sudan University of Science and Technology.

Heidaria, S., Sherkata, E. and Moosavia, S. 2014. A Novel Closed-Domain Question Answering System Fully Implemented Based on GATE Framework.

Ibn Baz, Abdul Aziz and AlUthaymeen, Muhammad and Al-Madkhalee, R. 2003. *Three Essays on the Obligation of Veiling. edited by Translated by Abu Maryam Ismaeel Alarcon Toronto: Al,*. Ibaanah Book Publishing.

Keleg, A. and Magdy, W. 2022. SMASH at Qur'an QA 2022: Creating Better Faithful Data Splits for Low-resourced Question Answering Scenarios *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Malhas, R., Mansour, W. and Elsayed, T. 2022. Qur'an QA 2022: Overview of the first shared task on question answering over the holy qur'an *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Maraoui, H., Haddar, K. and Romary, L. 2021. Arabic factoid Question-Answering system for Islamic sciences using normalized corpora. *Procedia Computer Science*. **192**, pp.69–79.

Mellah, Y., Touahri, I., Kaddari, Z., Haja, Z., Berrich, J. and Bouchentouf, T. 2022. LARSA22 at Qur'an QA 2022: Text-to-Text Transformer for Finding Answers to Questions from Qur'an *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Mostafa, A. and Mohamed, O. 2022. GOF at Qur'an QA 2022: Towards an Efficient Question Answering For The Holy Qu'ran In The Arabic Language Using Deep Learning Based Approach *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Premasiri, D., Ranasinghe, T., Zaghouani, W. and Mitkov, R. 2022. DTW at Qur'an QA 2022: Utilising Transfer Learning with Transformers for Question Answering in a Low-resource Domain *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Saeedi, P., Heidari, S. and Farhoodi, M. 2014a. Creating quranic question taxonomy *In*: *2014 22nd Iranian Conference on Electrical Engineering (ICEE)*. IEEE, pp.1070–1074.

Saeedi, P., Heidari, S. and Farhoodi, M. 2014b. Creating quranic question taxonomy *In*: *2014 22nd Iranian Conference on Electrical Engineering (ICEE)*., pp.1070–1074.

Sheker, M., Saad, S., Abood, R. and Shakir, M. 2016a. Domain-specific ontology-based approach for Arabic question answering. *Journal of Theoretical and Applied Information Technology*. **83**(1), p.43.

Sheker, M., Saad, S., Abood, R. and Shakir, M. 2016b. Domain-specific ontology-based approach for Arabic question answering. *Journal of Theoretical and Applied Information Technology*. **83**(1), p.43.

Singh, N. 2022. niksss at Qur'an QA 2022: A Heavily Optimized BERT Based Model for Answering Questions from the Holy Qu'ran *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.

Wasfey, A., Elrefai, E., Marwa, M. and Haq, N. 2022. Stars at Qur'an QA 2022: Building Automatic Extractive Question Answering Systems for the Holy Qur'an with Transformer Models and Releasing a New Dataset *In*: *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022)*.