Original Research Paper

A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification

Suhaib Kh. Hamed and Mohd Juzaiddin Ab Aziz

Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology,
University Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia

Corresponding Author:
Suhaib Kh. Hamed
Center for Artificial
Intelligence Technology
(CAIT), Faculty of Information
Science and Technology,
University Kebangsaan
Malaysia, Bangi, 43600,
Selangor, Malaysia
Tel:0060-1139044355
Email: Suhaib83.programmer@gmail.com

Abstract: In spite of great efforts that have been made to present systems that support the user's need of the answers from the Holy Quran, the current systems of English translation of Quran still need to do more investigation in order to develop the process of retrieving the accurate verse based on user's question. The Islamic terms are different from one document to another and might be undefined for the user. Thus, the need emerged for a Question Answering System (QAS) that retrieves the exact verse based on a semantic search of the Holy Quran. The main objective of this research is to develop the efficiency of the information retrieval from the Holy Quran based on QAS and retrieving an accurate answer to the user's question through classifying the verses using the Neural Network (NN) technique depending on the purpose of the verses' contents, in order to match between questions and verses. This research has used the most popular English translation of the Quran of Abdullah Yusuf Ali as the data set. In that respect, the QAS will tackle these problems by expanding the question, using WordNet and benefitting from the collection of Islamic terms in order to avoid differences in the terms of translations and question. In addition, this QAS classifies the Al-Baqarah surah into two classes, which are Fasting and Pilgrimage based on the NN classifier, to reduce the retrieval of irrelevant verses since the user's questions are asking for Fasting and Pilgrimage. Hence, this QAS retrieves the relevant verses to the question based on the N-gram technique, then ranking the retrieved verses based on the highest score of similarity to satisfy the desire of the user. According to F-measure, the evaluation of classification by using NN has shown an approximately 90% level and the evaluation of the proposed approach of this research based on the entire QAS has shown an approximately 87% level. This demonstrates that the QAS succeeded in providing a promising outcome in this critical field.

Keywords: Holy Quran, Question Answering System, Neural Network Classification, Question Expansion Technique

Introduction

With the global growing demand for Islamic knowledge by both Muslims and Non-Muslims, which is based on the Holy Quran as the major source of knowledge, law, conduct and wisdom, most of the systems that facilitate the search for the contents of the Quran remain as a significant challenge. In spite of there have been tools using for searching on Quran in recent years; most of these tools are using based on a keyword search, which means that the users need to know the exact keywords before starting the process of searching on the Holy Quran. In general, the existing search engines still face problems such as word mismatch or retrieve many irrelevant documents, particularly when the user's queries are not specific enough (Imran and Sharan, 2009; Ishkewy and Harb, 2015).

In the recent years, Question Answering System (QAS) has been investigated extensively and automatic QAS has become an interesting research field and the

results have shown an obvious improvement in its performance. In particular, during the last decade, a number of Question Answering Systems have emerged, which has been mainly driven by the Text REtrieval Conference (TREC). The Text Retrieval Conference (TREC) is developed for retrieving accurate information based on specific fields, for example classifies the information of academic fields in order to retrieve exact documents or information of the proposed questions.

One of the functions of the QA System is to processes and analyzes the users' questions efficiently in order to retrieve accurate answers to satisfy the users' needs of information. These systems use unrestricted text as a primary source of knowledge Sundblad (2007). There are two important factors to ensure the success of QAS; (1) analyze the users' needs (queries) efficiently using Natural Processing Language (NLP) and (2) classify and manage the documents that contain the candidates answers accurately based on document classification phase. Therefore, the accurate matching between users' questions and the proposed answers will be found effectively (Harb *et al*., 2009; Tan *et al*., 2009).

The Holy Quran as a book is not classified on subjects and its verses describe many topics and many verses even from different chapters converge within the same topic. The number of verses and chapters may share similar topics such as Faith and Morality. One of the solutions to tackle this issue is the Quran classification; therefore, there is an essential need for the Quran classification to classify the Quran based on its content for better information management and to determine the passage of verses that contain relevant verses to the user's query. Since this research dealing with a sacred script, therefore, the classification must be context sensitive. Moreover, there is a need to comprehend the current and possible classifications for the Quranic verses (Al-Kabi *et al*., 2013). The Holy Quran is regarded as a text is complex in its structure and also diversified in its styles of expression (Nassimi, 2008). Artificial Neural Network (ANN) classifier has been effectively applied in many fields and has shown a promising performance in documents classification due to its ability to recognize the complex patterns existing in the data (Ramlall, 2009; Mohammed and Omar, 2012; Patra and Singh, 2013).

In that respect, the QA systems should be able to readily and efficiently search for documents or passages relevant to the user's query, in order to retrieve the intended answers. Moreover, these systems need to locate the range or the passage of these answers and select the best of them or the most relevant to provide it to the user. For building an effective QAS that allows a user to ask a question in natural language and receive the answer briefly, the more advanced technologies are required that often combine related techniques of more established tasks; Information Retrieval (IR), Information Extraction (IE) and Natural Language Processing (NLP) (Aunimo, 2007; Hu, 2006). However, the general structure of QAS consists of three main phases which are: (1) Question Analysis, (2) Document Retrieval and (3) Answer Extraction (Hirschman and Gaizauskas, 2001; Kwok *et al*., 2001).

Related Work

Several studies have presented to facilitate and develop the search process in the Holy Quran and on how to build the Islamic ontology, although, most of these studies were directed to the Arabic user, it has presented studies investigate in the Quran translations for other languages. Ishkewy and Harb (2015) Produced an Islamic ontology that includes the main concepts of the Islamic domain such as Hadith and Tafseer (interpretation of Quran). They designed the Islamic Semantic Web Search Engine (ISWSE) that searches in the Holy Quran based on the Islamic ontology.

In that respect, Gusmita *et al*. (2014) presented research in the development of the Question Answering System (QAS) based on a combination of two approaches. The first one based on relevant documents and the second one based on a rule-based method, which this combination is focusing on answer extraction from Indonesian translation of Quran. The system performance is still unable to increase the accuracy in delivering correct answers. Similarly, Abdelnasser *et al*. (2014) proposed a Question Answering System on the Holy Quran that receives a question in the Arabic language as an input and retrieves semantically relevant verses as a passage that is probably to contain the answer by using the Quranic ontology. As well as, they have presented a new taxonomy for the Quranic Named Entities and constructed an Arabic Question Classifier.

In conjunction with that, Yauri *et al*. (2013) presented a semantic search system for Quranic knowledge by using an ontology assertion capability. They have used the existing ontology of the Leeds University, to link the concepts in the Holy Quran with various relations that exist between these concepts to semantically retrieve verses in according to user's query. This system has shown noteworthy results with respect to effectively retrieve the Quranic knowledge. At the same time, Yahya *et al*. (2013) have presented a semantic search in the Holy Quran based on Cross-Language Information Retrieval (CLIR). They produced a bilingual ontology for the Holy Quran made up of concepts based on Quranic Arabic corpus ontology that created by Dukes (2015) and they realized that the most of the documents are belonging to the main concept whereas others documents are not belonging to any of these concepts in English translation. With regard to Malay translation, the result is better than the English translation.

In the meantime, Khan *et al*. (2013) have developed a simple ontology for the Holy Quran, which includes the words that refer to animals that are mentioned in the Quran in order to enhance the semantic search process in

Quran. This ontology was implemented by using protégé editor and they used SPARQL query language to retrieve the answers based on the user's query. This ontology presents 167 direct or indirect references to the animals in the Quran. The relation that used in this ontology is taxonomy. They mention that the concepts of the semantic web could be used for the implementations of semantic search in the Holy Quran. They have stated that the WordNet is a significant structure that gives a dynamic touch to information retrieval from documents as well as web pages.

In the same vein, Shoaib *et al.* (2009) have proposed a model able to perform a semantic search. The research aim is to address the deficiencies of keywords search and the issues that related to semantic search on the Holy Quran. This model exploits the relations of WordNet in a relational database model. This model has been implemented on Surah Al-Baqrah in latest tools. The precision of this model prototype is far better than a traditional keyword search. In contrast, Saad *et al.* (2008) have presented an ontological work to extract keywords and key phrase candidate for developing the ontology of Islamic literature. They produced an algorithm for automatic extraction of the keywords. They proposed a general and skeletal methodology and lifecycle for creating the ontology of Islamic literature. In addition, they applied their approach to English text for mining ontologies from natural language.

Materials and Methods

Data Set

Since there are several translations of Holy Quran, depending on the understanding of the translator for the Holy Quran and his style. Therefore, this research has used the most popular translation of Quran among others translations that covers a large number of readers of the Holy Quran in the English language which is the English translation of the Holy Quran of Abdullah Yusuf Ali (YA) (2003) as the reference data set. This research investigates the questions that refer to identified named entities, which are the two pillars of Islam Pillars: Fasting and Pilgrimage. Since the "Al-Baqarah" chapter is the longest chapter of the Holy Quran, as well as, Al-Baqarah surah contains the largest number of verses that are talking about the Fasting topic which is 5 verses and 10 verses talking about the Pilgrimage topic. Therefore, this research has used "Al-Baqarah" Surah, as a data set that represents the sample of the Holy Quran.

Methodology

This study has proposed a Question Answering System (QAS) based on the semantic search that exploits the WordNet synonyms and a collection of Islamic terms that created by this research to expand the question. The Neural Network classifier has been employed to classify the Al-Baqarah Surah into Fasting and Pilgrimage verses to specify the candidate passage and by using the N-gram technique, the QAS will retrieve a set of relevant verses based on user's query. The verses will be ranked based on scoring matched words function to provide the user with correct verses based on its question. The Fig. 1 illustrates the general architecture of this QAS, which basically consists of three parts, which are: Question Analysis, Document Retrieval and Answer Selection.

Question Answering System

For building an effective QAS that allows a user to ask a question in natural language and receive the answer briefly, the more advanced technologies are required that often combine related techniques of more established tasks; Information Retrieval (IR), Information Extraction (IE) and Natural Language Processing (NLP) (Aunimo, 2007; Hu, 2006). Therefore, to build effective QAS, many of the phases and tasks should be involved within the QAS to configure this integrated system and these main and sub-phases as previously illustrated in Fig. 1 as following.

Question Analysis Module

The first module in the QAS is the Question Analysis, which is considered as a one of the important parts of the QAS that consists of two phases, which are question pre-processing and question expansion. According to (Karyawati *et al.*, 2015), there are two main procedures for any QAS, the first procedure is analyzing the structure of user's question efficiently and the second procedure is transforming the question into a meaningful question formula that compatible with QAS's domain.

Question Pre-Processing

The first phase in the question analysis module is the question pre- processing which is considered as one of the important processes that responsible for getting rid of the punctuation marks or words that are redundant in the computational analysis and do not have any worthy value in the searching process by using Normalization and Stop Words Removal technique. It is important to specify the significant words that considered as valuable and dismiss the words that do not contribute to differentiating between the documents (Ramasubramanian and Ramya, 2013). As well as in this phase removing various affixes from words to reduce the number of different words that have the same root, to have exactly matching stems and retrieve all the verses that have this stem, which is most likely related to the user's need, this achieved by using a Stemming technique.
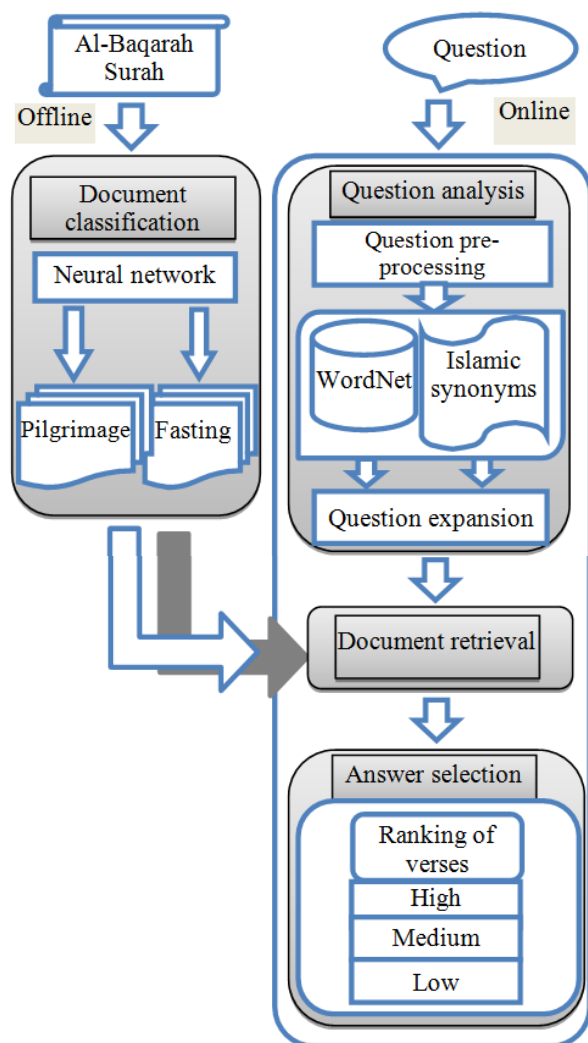
Fig. 1. Question answering system architecture

Question pre-processing phase		
User Question	What did God reveal in the month of Ramadan?	
Pre-processing Phase	Normalization	?
		what did god reveal in the month of ramadan
	Stop Words Removal	what, did, in, the, of
		god reveal month ramadan
	Stemming	god reveal month ramadan
		god reveal month ramadan

Fig. 2. Question pre-processing phase

The level of efficiency at this phase will affect the performance accuracy in the later phases. This phase could be divided into three sub-processes, Normalization, Stop Word Removal and Stemming, which run sequentially, as it is shown in the Fig. 2.

Question Expansion

In this phase, the system will expand the user question and generate a number of questions by taking advantage of the synonyms of WordNet database as well as the collection of Islamic terms that created by the researcher. In order to find the synonyms for each word in the user's query to cover all possible meanings that might be used in the translation of the Holy Quran, such as illustrated in the Table 1. To ensure the matching will happen between the user's words and the words of the Quran. When the users type their queries, the system tries to involve the ontology knowledge to improve the Query Expansion in order to enhance the probability of relevancy (Wang *et al*., 2012). This research has used a combination of WordNet synonyms and the Islamic synonyms. The collection of Islamic terms has been collected from the many English translations of the Holy Quran, Hadith and Tafseer (Interpretation of Quran) that particularly related to the themes of Fasting and Pilgrimage. These Islamic terms collected by the researcher and has approved by experts.

Thus, the question expansion is performed through generating a number of questions as it is illustrated in Table 2, which depends on the number of retrieved synonyms from the WordNet and the collection of Islamic terms; therefore, the process of generating the questions by QAS takes all the probabilities to generate these questions from these synonyms.

Document Retrieval Module based on Document Classification

The second module in the QAS is the Document Retrieval. The function of the document retrieval module is not to find actual verses to the user's query, but to specify verses that are likely to contain an answer. The main aim of document retrieval is extracting relevant verses from the Holy Quran, before sending them to the next phase which is the answer selection module. This module includes verses classification through Document Classification phase.

Since the research scope focuses on the users' questions that refer to the two pillars of Islam: Fasting and Pilgrimage. Therefore, Al-Baqarah Surah will be classified into two classes Fasting and Pilgrimage. The main goal of text classification is to reduce the searching space by identifying the passages of information that are relevant to the particular topic (Baharudin *et al*., 2010). Consequently, map user questions to their corresponding verses, because there are a limited number of possible answers (verses).

Table 1. Synonyms of WordNet and the collection of Islamic terms

Terms	Source	Synonyms
God	WordNet	God, supreme being, deity, divinity, god, immortal, idol, graven image
	Collection of Islamic Terms	Allah, almighty
Reveal	WordNet	Reveal: Bring out, unveil, unwrap, disclose, let on, bring out, discover, expose, divulge, break, give away, let out, uncover
	Collection of Islamic Terms	-
Month	WordNet	Calendar month
	Collection of Islamic Terms	-
Ramadan	WordNet	-
	Collection of Islamic Terms	Ramadhan, Ramazan

Table 2. Expanding questions

Generating questions
Q: God reveal month ramadan
Q1: God bring out month ramadan
Q2: God reveal calendar month ramadan
Q3: Allah reveal month ramadan
Q4: God reveal month ramazan
Q n: Allah reveal month ramadhan
(n indicates to the number of expanding questions based on a number of terms and its synonyms)

This research has used the Artificial Neural Network (ANN) to classify the verses of Al-Baqarah Surah. ANN has been effectively applied in many areas of artificial intelligence, for example, NLP, pattern recognition and classification tasks (Mohammed and Omar, 2012).

This research has used the WEKA toolkit to implement the classification of Al-Baqarah surah based on NN classifier. WEKA is a machine learning platform that contains a collection of the popular machine learning algorithms that could be used for practical data mining and machine learning applications, As well as it includes many tools for data pre-processing (Witten and Frank, 2005). Therefore, this research highlights the tasks that carried out in WEKA with regard to the classification of the Al-Baqarah Surah, which consists of the following steps: A Training set and Filtered classifier that combines filter and Neural Network classifier.

Training Set

The training set consists of the list of 150 instances of verses sharing a set of attributes, where this training set represents 80% of the total number of the data set, which means that the 20% of the remaining data will be for the testing set. The training set has three classes, 50 examples for the Fasting class, 50 examples for the Pilgrimage class and 50 examples for none class that are not related to any of these mentioned classes. Each example labeled by value represents its class. This phase is considered the most important phase because it is through these examples, the Neural Network classifier could learn and then able to predict the classes of verses easily and hence, to build the best classifier mainly depends on the quality of these examples.

Filtered Classifier

The Filtered Classifier is a combination of the String to Word Vector filter and Neural Network classifier based on Back- Propagation Network. This filter that could be able to deal with string attributes directly, without the need for the filter in an isolated stage to process and transform the verses. In WEKA, the task of the filter is similar to the task of data pre-processing phase, which includes several tasks that use to pre-process the data. The raw verses are data and firstly should transform into a form appropriate for learning by generating a dictionary of terms from all these verses in the training set and assign a numeric attribute for each term using the filter String to Word Vector. A word vector is a numeric vector representing the values is derived from the number of occurrences of each word in the verse.

The function of The Back-propagation Neural Network (BPNN) classifier is to compare the weights of all feature sets that extracted from the Quranic verses with the weights of predefined classes based on the training set to determine each verse to its class. BPNN classifier is a network of units consisting of the input layer, the hidden layer and the output layer. All the neurons in the hidden layer and the output layer have biases, which are a connection from units whose activation function is always 1, the function of bias is similar to the weights. There are two main phases of the back-propagation learning process, the forward phase and the backward phase. In the forward phase, the input signals transfer forward through the network layer by layer and finally generating the actual output of the network. The actual output that generated is compared with the desired output, if there is a difference, this clearly indicates that there is an error; in order to calculate the error and reduce it, the error signals will be generated and then propagated in a backward direction. In the backward phase, small adjustments should be performed in weights of the network to reduce the sum squared errors. Back-propagation learning has been implemented successfully to solve many difficult problems (Aljawfi *et al.*, 2014). The algorithm of BPNN could be summarized in these steps:

- Back propagation has initial weights (random), normally in the range [-0.5, 0.5]
- Update the weights to achieve output consistent with the training sets
- Compute the error as desired output minus actual output
- Error e = Y desired -Y actual
- The weights need to be adjusted to decreasing the error

The Neural Network will be more understanding and knowledgeable about its environment after each iteration of the learning process (Kaur, 2012). Moreover, the learning process by training the algorithm of a Neural Network is not merely an issue of memorizing the mapping relations between the inputs and the outputs of the provided examples, but in fact, it is to extract the internal rules and distinctive features from these examples which are obscure to traditional user (De Houwer *et al.*, 2013). The settings of NN classifier is fixed such as, Learning Rate = 0.3, Momentum = 0.2, Hidden Layers= (Input+output)/2. The setting of the filter is outperformed based on important tasks such as Stop Word Removal, Stemming and Term Frequency (TF) transformation.

Answer Selection Module

The final part of QAS is the answer selection, the representation of the questions and the representation of the intended verses that are probable to contain the answer are matched against each other and a group of candidate verses is presented, ranked according to the likelihood of correctness and relevance. These intended verses that represent the verses of Fasting and Pilgrimage are retrieved based on document retrieval part and presented to the answer selection part. This research has used the N-gram technique to extract the answers and according to the ranking process based on the Words Matching Scoring function provides the user with the most relevant verse to the question. Figure 3 shows how to extract the answers based on the N-gram technique.

N-Gram Technique

This research has used the N-gram technique based on unigrams, bigrams, to retrieve the verses in terms of common n-gram between questions and verses. The n-gram is applied for each question, which divides the sequence of words based on the white space and then generates a list of words or segments of words based on the selected n-gram. Thus based on this list of generated words, the answer selection module will extract the similar words from the verses based on the selected n-gram. Based on the experiments conducted by this research, the selection of the n-gram size is important because a small n-gram size would cause many matchings, whereas, a large size would produce very few matches. N-grams have been successfully used in many text applications of language processing, which includes the identifying and measuring the text reuse in journalism (Adeel Nawab *et al.*, 2012).

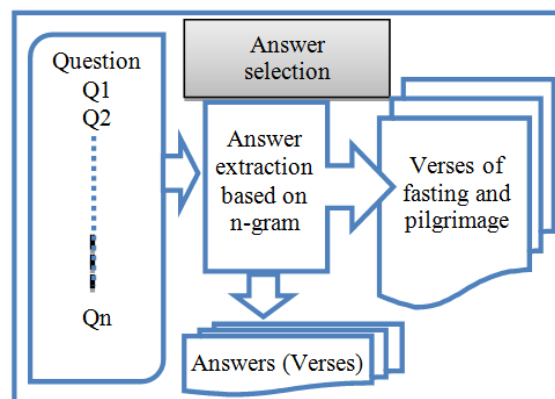Figure 4 illustrates how to extract the verses using the unigram and bigram based on the questions.

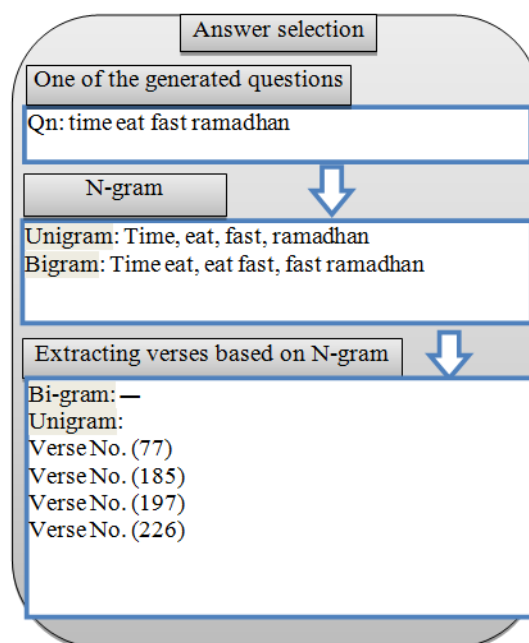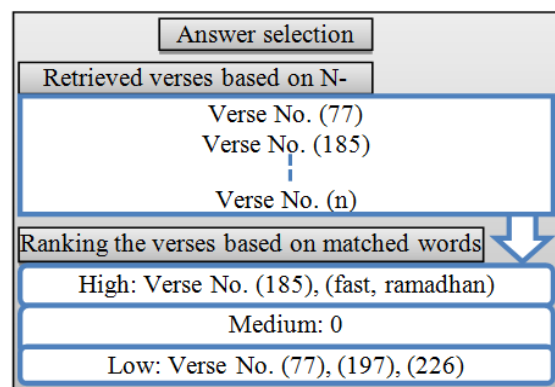Fig. 3. Answer selection module

Fig. 4. N-gram technique

Fig. 5. Verses ranking technique

Verses Ranking

In order to rank the verses, Words Matching Scoring function is applied for each of these verses that are retrieved based on N-gram to count the number of similar words between the expanded questions and these verses (Gusmita *et al*., 2014) and then ranking all these scored verses. Certainly, the relevant verse that related to the user's query contains the largest number of similar words with the user's question. Therefore, the verse with highest scored will be ranked at the high level and other verses will be ranked based on this function. If an answer contains the words of the user's question, it is probably this is the answer that the user is looking for it (Grappy *et al*., 2011). Figure 5 shows the how to rank the verses based on the scoring function.

Evaluation

The evaluation based on the F-score has been applied to four systems of the QAS. Recall and Precision are conventional metrics used for information retrieval systems. F-score is the harmonic mean of the Recall and Precision (Mathur and Haider, 2015). These systems are different in terms of using the classification method, or using the collection of Islamic terms or using the ranking technique. In regard to the evaluation of document classification, this research has measured the evaluation based on the F-score among four proposed classifiers, which are different from using the training set and the configuration of filter and classifier. The experts have assisted this research in determining, which retrieved answers (verses) are related to the user's query based on QAS, as well determine which verses that belong to the Fasting and Pilgrimage category based on NN classification.

Results

To discuss the results of the whole QA System, firstly, should review the results of document classification part. The document classification is considered as an important part of the whole QAS because it increases the accuracy of retrieval of the correct answer. The document classification based on NN has shown a high accuracy among all the tests conducted in this critical domain. The evaluation of NN classifier has shown a high value based on the F-score measure, which is 90% level. Based on the NN classification, all the verses of Fasting and Pilgrimage of Al-Baqarah surah are assigned to the correct class as it showed in Table 3 and reduced the irrelevant verses in each class and excluded 246 out of 286 verses that are not related to the user's query. According to the outcomes obtained from NN classifiers based on many tests, the use of Tafseer (Interpretation of Quran) instances without Quranic verses that related to these classes in the training set present a bad classification.
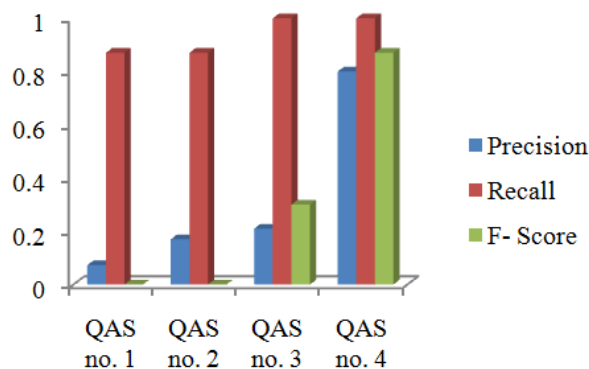

Fig. 6. The evaluations of four systems

Table 3. NN classification

Class	Truth fasting	Truth pilgrimage	Truth none	F-score
Predicated as fasting	5	0	0	0.90
Predicated as pilgrimage	0	10	0	
Predicated as none	12	13	246	

This due to the instances of Tafseer have explained clearly which have distinctive terms that refer to these classes. Therefore, this classification will increase the accuracy of retrieving relevant verses and reduce the search time and memory used in the implementation of QAS.

This research highlights on four significant experiments of QA Systems in terms of improvements to the proposed QAS in order to raise the accuracy of performance of this QAS. As shown in the Fig. 6, the stages of the evolution of this QAS in terms of accuracy of the retrieval of the required verse to the user's question, where these QA Systems were tested based on the questions proposed by the experts. According to the values of the evaluation based on Precision, Recall and F-score, the first QAS has not used the NN classification, wherein this experiment; the whole Al-Baqarah Surah has been processed, the results showed a poor value for the accuracy of the QAS. This refers to many irrelevant verses were retrieved due to this system did not use the document classification based on NN for the Al-Baqarah Surah depending on the domain of user's search. The advantage of Quran classification is reducing the irrelevant verses depending on the user's question domain. In the second QAS, the NN classifier was used to classify Al-Baqarah Surah based on the domain of the user's search. Thus, the value of Precision has increased slightly. The Recall values of the first and the second QAS still poor and refer to the weakness of retrieving process of the correct verses to the user's question and this due to non-use of the collection of Islamic terms in the first and second experiments, which lead to mismatches between the vocabulary of the question and the Quran. In addition to, the WordNet synonyms used in these experiments do not cover all the Islamic terms.

The Precision value of the third QAS experiment was increased slightly from what it was with a noting that the Recall value became very high and that is due to the use of Islamic terms with WordNet together, wherein all the previous experiments, the retrieved verses were based on N-gram technique. The final QAS which is proposed by this research showed a high value of evaluation based on the F-score which is an approximately 87% level depending on the results of Precision and Recall as it illustrated in Fig. 6. This is due to the use of ranking process based scoring function in this experiment of QAS to select the most relevant verse based on the user's question from the retrieved verses based on the N-gram.

Conclusion

This research provided an integrated QA system to retrieve the accurate answer from the Holy Quran according to the user's question, where this QAS differs from other systems that rely on the keywords search, in terms of might not retrieve any answer to the user's question, or retrieve an incorrect answer, or several answers, including the correct answer. Therefore, this QAS is deal with several problems such as, there are many translations of the Holy Quran and each translation has owned vocabularies may differ from the others, or these translations may use Islamic vocabularies that are unfamiliar to the speakers of English language. Thus, this QAS has used a collection of Islamic terms to expand the user's question to ensure the matching will happen between the user's words and the words of the Quran. In addition to, the Quran has a large number of topics and the user's question directed to the topics of Fasting and Pilgrimage. Therefore, it was classified the Quran that is represented by Al-Baqarah surah to the classes of Fasting and Pilgrimage. Consequently, it will be excluded the verses that are irrelevant to the scope of user's search. Therefore, this QAS will retrieve a number of verses that are relevant to the user's question and by using a ranking process, this system will provide the user with accurate verse according to its question based on the larger number of similar words between the question and the Quranic verses. Although it has proven that this research has an achieved its goal, but could be in the future develop this research by building an Islamic lexical database of English terms that contains the terms that have used in the English translations of the Holy Quran to include the entire Holy Quran.

Acknowledgement

Author's Contributions

Both the authors have equally contributed to this manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Abdelnasser, H., R. Mohamed, M. Ragab, A. Mohamed and B. Farouk *et al*., 2014. Al-Bayan: An Arabic question answering system for the Holy Quran. Proceedings of the EMNLP Workshop on Arabic Natural Langauge Processing, (NLP' 14), Doha, Qatar, pp: 57-64. DOI: 10.3115/v1/w14-3607

Adeel Nawab, R.M., M. Stevenson and P. Clough, 2012. Detecting text reuse with modified and weighted n-grams. Proceedings of the 1st Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the Main Conference and the Shared Task and Volume 2: Proceedings of the 6th International Workshop on Semantic Evaluation, (WSE' 12), Stroudsburg, PA, USA, pp: 54-58.

Aljawfi, O.M., N.M. Nawi and N.A. Hamid, 2014. Enhancing back propagation neural network with second order con-jugate gradient method for fast convergence. Proceedings of the 1st International Conference of Recent Trends in Information and Communication Technologies, (ICT' 14), pp: 575-586.

Al-Kabi, M.N., B.M.A. Ata, H.A. Wahsheh and I.M. Alsmadi, 2013. A topical classification of Quranic Arabic text. Proceedings of the Taibah University International Conference on Advances in Information Technology for the Holy Quran and its Sciences, Dec. 22-25, Madinah, Saudi Arabia, pp: 252-257.

Aunimo, L., 2007. Methods for Answer Extraction in Textual Question Answering. 1st Edn., University of Helsinki, ISBN-10: 9521039922, pp: 127.

Baharudin, B., L.H. Lee and K. Khan, 2010. A review of machine learning algorithms for text-documents classification. J. Adv. Inform. Technol., 1: 4-20. DOI: 10.4304/jait.1.1.4-20

De Houwer, J., D. Barnes-Holmes and A. Moors, 2013. What is learning? On the nature and merits of a functional definition of learning. Psychonomic Bull. Rev., 20: 631-642. DOI: 10.3758/s13423-013-0386-3

Dukes, K., 2015. Statistical parsing by machine learning from a classical Arabic Treebank. The University of Leeds.

Grappy, A., B. Grau, M.H. Falco, A.L. Ligozat and I. Robba *et al.*, 2011. Selecting answers to questions from Web documents by a robust validation process. Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Aug. 22-27, IEEE Xplore Press, Lyon, pp: 55-62. DOI: 10.1109/WI-IAT.2011.210

Gusmita, R.H., Y. Durachman, S. Harun, A.F. Firmansyah and H.T. Sukmana *et al.*, 2014. A rule-based question answering system on relevant documents of Indonesian Quran Translation. Proceedings of the International Conference on Cyber and IT Service Management, Nov. 3-6, IEEE Xplore Press, South Tangerang, pp: 104-107. DOI: 10.1109/CITSM.2014.7042185

Harb, A., M. Beigbeder and J.J. Girardot, 2009. Evaluation of question classification systems using differing features. Proceedings of the International Conference for Internet Technology and Secured Transactions, IEEE Xplore Press, London, pp: 1-6. DOI: 10.1109/ICITST.2009.5402567

Hirschman, L. and R. Gaizauskas, 2001. Natural language question answering: The view from here. Natural Lang. Eng., **7**: 275-300. DOI: 10.1017/S1351324901002807

Hu, H., 2006. A study on question answering system using integrated retrieval method. PhD. Thesis, The University of Tokushima.

Imran, H. and A. Sharan, 2009. Thesaurus and query expansion. Int. J. Comput. Sci. Inform. Technol., 1: 89-97.

Ishkewy, H. and H. Harb, 2015. ISWSE: Islamic Semantic web search engine. Int. J. Comput. Applic., 112: 37-43. DOI: 10.5120/19664-1337

Karyawati, A.E., E. Winarko, A. Azhari and A. Harjoko, 2015. Ontology-based why-question analysis using lexico-syntactic patterns. Int. J. Electrical Comput. Eng., 5: 318-332.

Kaur, T., 2012. Implementation of backpropagation algorithm: A neural net-work approach for pattern recognition. Int. J. Eng. Res. Develop., 1: 30-37.

Khan, H.U., S.M. Saqlain, M. Shoaib and M. Sher, 2013. Ontology based semantic search in Holy Quran. Int. J. Future Comput. Commun., 2: 570-575.

Kwok, C., O. Etzioni and D.S. Weld, 2001. Scaling question answering to the web. ACM Trans. Inform. Syst., 19: 242-262. DOI: 10.1145/502115.502117

Mathur, A. and M. Haider, 2015. Question answering system: A survey. Proceedings of the International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, May 6-8, IEEE Xplore Press, Chennai, pp: 47-57. DOI: 10.1109/ICSTM.2015.7225389

Mohammed, N.F. and N. Omar, 2012. Arabic named entity recognition using artificial neural network. J. Comput. Sci., 8: 1285-1293.

Nassimi, D.M., 2008. A thematic comparative review of some English translations of the Qur'an. Ph.D. Thesis, University of Birmingham.

Patra, A. and D. Singh, 2013. A survey report on text classification with different term weighing methods and comparison between classification algorithms. Int. J. Comput. Applic., 75: 14-18. DOI: 10.5120/13122-0472

Ramasubramanian, C. and R. Ramya, 2013. Effective pre-processing activities in text mining using improved porter's stemming algorithm. Int. J. Adv. Res. Comput. Commun. Eng., 2: 2278-1021.

Ramlall, I., 2009. Artificial intelligence: Neural networks simplified. Int. Res. J. Finance Econom. Forthcom.

Saad, S., N. Salim and N. Omar, 2008. Keyphrase extraction for Islamic Knowledge ontology. Proceedings of International Symposium on the Information Technology, Aug. 26-28, IEEE Xplore Press, Kuala Lumpur, Malaysia, pp: 1-6. DOI: 10.1109/ITSIM.2008.4631711

Shoaib, M., M.N. Yasin, U. Hikmat, M.I. Saeed and M.S.H. Khiyal, 2009. Relational WordNet model for semantic search in Holy Quran. Proceedings of the International Conference on Emerging Technologies, Oct. 19-20, IEEE Xplore Press, Islamabad, pp: 29-34. DOI: 10.1109/ICET.2009.5353208

Sundblad, H., 2007. Question classification in question answering systems. Linköpings Universitet.

Tan, W., J. Cao and H. Li, 2009. Algorithm of shot detection based on SVM with modified kernel function. Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, Nov. 7-8, IEEE Xplore Press, Shanghai, pp: 11-14. DOI: 10.1109/AICI.2009.243

Wang, H., Y. Guo, X. Shi and F. Yang, 2012. Conceptual Representing of Documents and Query Expansion Based on Ontology. In: Web Information Systems and Mining, Wang, F.L., J. Lei, Z. Gong and X. Luo (Eds.), Springer, pp: 489-496.

Witten, I.H. and E. Frank, 2005. Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edn., Elsevier, Burlington, ISBN-10: 0080890369, pp: 664.

Yahya, Z., M.T. Abdullah, A. Azman and R.A. Kadir, 2013. Query translation using concepts similarity based on Quran ontology for cross-language information retrieval. J. Comput. Sci., 9: 889-897. DOI: 10.3844/jcssp.2013.889.897

Yauri, A.R., R. Abdul Kadir, A. Azman and M.A. Azmi Murad, 2013. Quranic verse extraction base on concepts using OWL-DL ontology. Res. J. Applied Sci. Eng. Technol., 6: 4492-4498.