International Journal of Advanced Trends in Computer Science and Engineering

Question Expansion Technique On The Different Translations Of The Holy Quran By Using WordNet And Islamic Synonyms

SuhaibKh. Hamed[1], **MohdJuzaiddin Ab Aziz** [2]
[1]Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, University Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia , Suhaib83.programmer@gmail.com
[2]Center for Software Technology and Management (SOFTAM), Faculty of Information Science and Technology,University Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia , juzaiddin@ukm.edu.my

ABSTRACT

The Holy Quran is the most significant religious text, which is followed by the believers of the Islamic religion. The translations of the Quran are the interpretation of its meaning in different languages, to assist those who are not familiar with the Arabic language.Most of the Quranic search systems that support the needs of who are non-native speakers of the Arabic language of the contents of the Holy Quran are based on keywords matching. Since there are several known English translations of the Holy Quran and each translation has its terminology based on the style of the translator. In addition to that, there are Islamic terms might be not used in the English language, such as "Hajj" or "Saum," where these terms might vary from one translation to another. Therefore, these search systems do not support the needs of the user, and often there is no matching between the terms of user's question and the terms of the Quran. Consequently, this research has used the synonyms of WordNet to expand the user's question, and to cover all the possibilities of the terms that might be used in the Quran translation, but there are many of Islamic terms are not available in the WordNet database. The WordNet has been enhanced with a set of Islamic synonyms that have been used in the well-known Islamic translations of the English language, which it was created by the researcher, to increase the likelihood of providing a sure matching between the question's words and the contents of the Quran translation. According to the results of this study, the evaluation of this research proved that the method of question expansion by using the both of the WordNet and the Islamic synonyms showed better accuracy and efficiency than the traditional keywords matching systems, or the system that uses only the WordNet synonyms in the question expansion method in terms of matching all user's questions words with documents and retrieve the correct answers to these questions.

Key words: Holy Quran, Question Expansion Technique, WordNet, Islamic Synonyms.

1. INTRODUCTION

Question expansion is a technique to extend the words of the question by adding new words that are not included in the original question similar to them in meaning in order to improve the quality of the retrieved documents in terms of relevance [1]. If there are few answers or even no answer is retrieved, an additional step by expanding the question which will lead to extend the search space and hence retrieving more candidates of answers. The process of expanding the posed question will retrieve a large collection of documents that possibly including the answers. The main goal of question expansion is to strengthen the question with relevant additional words such as synonyms to improve the precision of the search process, and hence to improve the recall of the retrieval [2].The availability of the Quran translations making the accessibility of written knowledge that related to Quran becomes faster and less complicated, especially for those who are not familiar with the Arabic language[3]. Muslims and non-Muslims, who are non-native speakers of the Arabic language, could be able to understand the Holy Quran through using its translations that provide them with the knowledge to comprehend a view of the Islamic religion [4]. These translations of the Quran are the interpretation of its meaning in different languages, to assist the readers who are not familiar with the Arabic language, to understand God's guidance and will. Muslim scholars commonly agree that it is big challenging to transfer the original Quran word by word into another language [5]. The translation of the Holy Quran is considered as a critical and controversial issue in the Islamic theology, due to several reasons, such as the words in Arabic could have multiple meanings based on the context. In addition, most of the Shariah scholars considered the translation of the Quran as the interpretation of the Quran [6].The most of the available search engines and current retrieval methods with regard to the Quran lack the semantic information retrievalthat are mostly based on keywords matching approach. In other words the users must define the proper keywords before beginning the operation of searching inside the verses of the Holy Quran[7]. Moreover, the Quranic keyword-based search methods use limited Islamic resources regarding with the Quran [8]. However, keyword based search without any semantic search retrieves many irrelevant verses, therefore, the users should take the trouble of searching for the answers inside each verse [9]. Generally, the current search engines still subject to problems, such as the most common problem is word mismatch, or another problem which is, retrieve many irrelevant documents, in particular, when the user's question is not clearly defined and specific enough [10]; [11].In its original Arabic language, the Quran is the comprehensible utterance of God for Muslims. However, its translation to another language has usually depended on the understanding of the human translators [12]. Consequently, each translation of the Holy Quran has its

Islamic terms, which differ from translation to another. For example, (pilgrimage, haj, hajj) all these terms refer to the pilgrimage. Therefore, the semantic search should be able to retrieve the related verses from the Holy Quran by using many techniques such as reformulating or expanding the question through using synonyms or using ontological resources as WordNet or thesaurus.A potential problem with any Information Retrieval system, especially with regard tothe Holy Quran, is that there is no matching between the user's question terms and the terms of documents, due to several reasons such as (i) The users may use terms differ from the formal terms that used in the documents. (ii) Most of the Information Retrieval studies have focused on the user's question extraction, but have not explored much on the correct mapping of the terms in the question with the ontology. Thus, these reasons will minimize the opportunities to retrieve accurate answers to their questions [13]. Moreover, the Islamic terms are different from a document to another and sometimes translated as it pronounced in the Arabic language, in addition, some of these Islamic terms have a mixture of synonyms that some are included in the English language, and some are not [14].

2. ONTOLOGY-BASED QUESTION EXPANSION TECHNIQUE

Ontologies are increasingly being used in the Information Retrieval domain as knowledge to support the semantic search [15]; [16]. When the user type its questions, the system tries to involve the knowledge of the ontology to enhance the question expansion technique in order to increase the probability of relevancy[17]. The ontology includes the definitions of the basic terms and the relations comprising the vocabularies of a specific field. In addition, it contains the rules of the combination of these terms and their relations to describe the extension vocabularies [18].Hu, Wang [19]have stated thatthere are various ontologies include different details at varying levels. These ontologies were categorizedregarding the level of generality:
1. Generic ontologies, which include general concepts, that are independent of any specific domains;
2. Domain ontologies, which include the concepts of a specific domain such as, geology, medicine or biology;
3. Application ontologies, which include the concepts that are specified for specific applications.

Karyawati, Winarko [20]have mentioned that the differences in systems approach in natural languages processing force the researchers to search for the appropriate ontology and questions representation to provide accurate results based on the nature of the QA System's domain. Allam, Haggag[21] pointed out to the presence of ontology is significant, and that could be used to expand the keywords of the question, which is WordNet ontology. The WordNet ontology is used in finding the synonyms of the question's keywords that could be used in matching with the words of the documents that might contain the answer phrase.

3. RELATED WORK

Several research have been presented to improve the question expansion technique on the Holy Quran. In spite of that most of these research were dedicated to the Arabic users, several studies have been presented that investigate into the question expansion technique for the Quran translations.Khan, Saqlain [22]have proposed a small English ontology based on a domain-specific ontology that specified for the birds and animals that mentioned in the Holy Quran to support the semantic search on the English translation of Holy Quran. This specific ontology of animals was created by using the protégé ontology editor tool. SPARQL query language has been implemented to illustrate the function of the ontology properly, and to examine the efficiency of this ontology and the relations between the concepts, and then to support the semantic search and performing the semantic manipulation for the query, not only focusing on keywords. Khan, Saqlain [22]concluded that based upon different types of queries using SPARQL, the results showed a high ability of Inference by using this ontology and successfully answer the queries.Shoaib, Yasin [9] have proposed a model that is capable of performing a semantic search. The aim of this research is to address the insufficiency of the keywords search and the issues related to semantic search in the Holy Quran. This research takes advantage of the relations of WordNet in a relational database model.This model has been executed on Al-Baqrah chapter in most recent tools.The precision of this model is more effectively than traditional keyword search. In contrast, Saad, Salim [22] have published a research on ontology about extracting the keywords and key phrases desired to develop the Islamic literature ontology. They created an algorithm for automatically keywords extracting. They suggest an overall and structural methodology as well as to its life cycle to construct the Islamic literature ontology. Furthermore, they implemented their project onto English text to extract ontologies from natural language.In addition,Hammo, Sleit [23] have investigated the use of question expansion for searching in the Holy Quran in the absence of diacritics, where queries are automatically augmented with related terms extracted from a vowelized index by applying a stemmer and a thesaurus of semantic classes. They conducted a set of experiments on searching words from the Quran and they concluded that question expansion for searching Arabic text gives promising results and the efficiency could be improved better than before.

4. MATERIALS AND METHODS

4.1 Dataset

This research uses the English translation by Abdullah Yusuf Ali (YA) (2003) of the Holy Quran as the reference data set. This translation is considered as the most common edition of the English translation of the Holy Quran among others editions of translations which covers large numbers of readers of the Holy Quran in its English language translation. Since the Fasting and the Pilgrimage are the two pillars of the five pillars of Islam ,therefore, this study focuses on the question of the "Fasting" and the "Pilgrimage" of the Holy Quran.

4.2 Question Analysis Module

Question analysis module is responsible for extracting the main terms of the user question that have significant value in the search process and retrieve the relevant answers. In

addition, this module is responsible for expanding user's question by taking advantage of the ontological resources such as WordNet and collection of Islamic terms. Hence, will be generated more than one question depending on the list of suggested synonyms, in order to maximize the opportunities of finding the similarities between the terms of the user's question and Quran's contents. Therefore, this module will handle the weakness of queries' terms and update these terms to be compatible with the terms of the Holy Quran, which maximize the accuracy of retrieved verses. Figure 1 illustrates the process of question expansion technique.
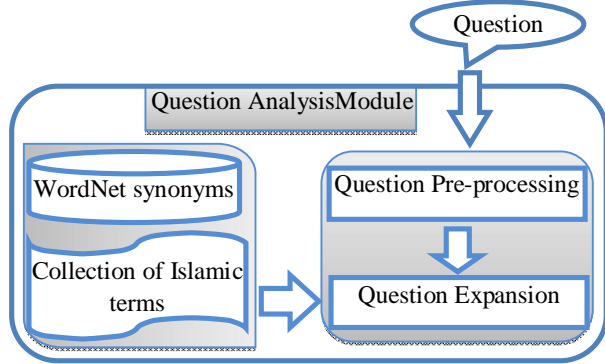

Figure 1 : Question analysis module

4.2.1 *Question Pre-processing*

The question preprocessing considered as the first phase in the question analysis unit which regards as a one of the significant processes that responsible for eliminating of redundant words in the searching process that do not have any worthy value in the computational analysis, as well as the punctuation marks by using Normalization and Stop Words Removal techniques.. Moreover, these words expand the process of retrieving irrelevant documents. The pre-processing phase is crucial in deciding the quality of the next phases. It is necessary to determine the important words that considered as valuable in the specific domain and exclude the words that do not lead to differentiation between the documents[24]. In addition, in this phase removing various suffixes from words to decrease the number of varied words that belong to the same root, in order to have a completely matching of stems and then retrieve all the verses which have the same stemthat is most probably represent the user's need, this operation will be performed by using a Stemming technique.The level of efficiency at this stage will impact on the accuracy of the performance in the next phases. This stage consists of three sub- procedures, Normalization, Stop Word Removal and Stemming which executed respectively. Figure 2 shows the whole process of Question Pre-processing.
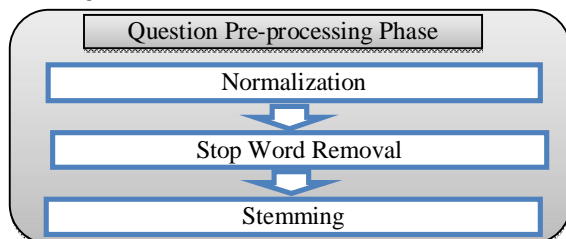

Figure 2: Question pre-processing phase

4.2.1 *Question Expansion Technique*

In this phase, the question expansion technique will expand the user's question and generate a number ofquestions by utilizing from the synonyms of WordNet database. In addition, the collection of Islamic terms and their synonyms, in order to find the synonyms for each word in the user's question to cover all possible meanings that might be used in the translation of the Holy Quran. The ontology plays a vital role in extracting the hidden information from the religious texts [25]. Hence, this system will achieve a precise matching between one or more of these questions and the Quranic verses. When the user types its question, the system tries to involve the knowledge of the ontology to enhance the question expansion in order to enhance the probability of relevancy[17].The question expansion technique was performedby generating a number of questions , which relies on the number of the retrieved synonyms from the both of the WordNet synonyms and the collection of Islamic synonyms; consequently, the QAS will generate the questions by take into account all the probabilities of generating these questions from these synonyms, as the following rules:

- If (the words in the question have synonyms in the WordNet and the Collection of Islamic Terms) Then
- Retrieve all these relevant synonyms
- Replacing these words with their synonyms and generating new questions
- End If

Thus, the main purpose of the question expansion technique is to extend the results scope and retrieve results more accurate of a given question in a semantically meaningful way. Figure 3shows the matching process of words and their synonyms. The large collection of synonyms could play important roles, especially when the knowledge in the ontology has multiple words representing the same term [14]. Therefore, the recall of the relevant information will increase.
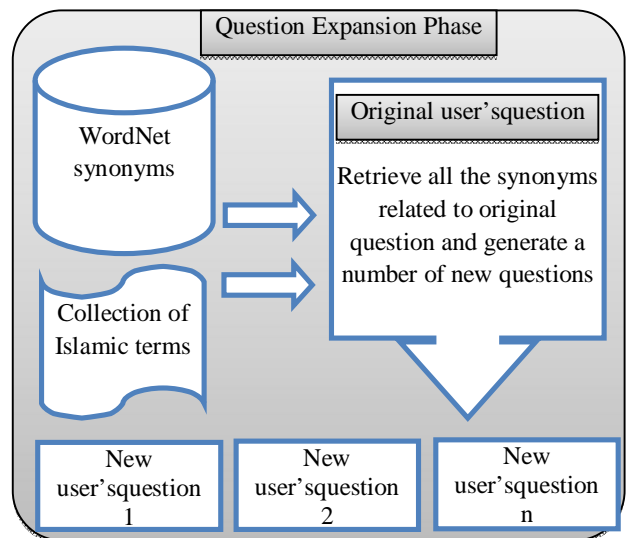

Figure 3: Question expansion technique

4.2.1.2 WordNet

The database of WordNet covers the vast majority of nouns, verbs, adjectives, and adverbs of the English language, and the relations among them. The WordNet is a significant database that gives a dynamic touch to information retrieval from documents through question expansion method[22]. The database of the WordNet is available for download from the project website[1]. This research has used WordNet, version 2.1. The following table 1 shows an example of words in the user's question and their synonyms from the WordNet.

Table 1: The WordNet synonyms

Word	Synonyms
time	clip, fourth dimension , sentence, clock time, prison term, metre
iftar	—
fast	fasting
ramadan	—

4.2.1.3 Collection of Islamic Terms

Since the Islamic terms are different from the translation to another and sometimes translated as it pronounced in the Arabic language such as Saum or Hajj, in addition, some of these Islamic terms have a collection of synonyms that some are included in the English language, and some are not, such as Fasting and Saum. All this diversity is due to the style of the translator and his perspective for the Holy Quran, which is originally written in the Arabic language. Thus, it is impossible to force the user to know which terms are used in this translation. Therefore, the question expansion technique has exploited the collection of Islamic synonyms that created by the researcher to cover all probabilities of synonyms that could be used in the English translation of Holy Quran, to provide accurate answers to the user's question. The researchers need an efficient mapping between the terms of the questions and the terms of system documents by providing an effective ontology for the proposed domain [20]. This study uses a combination of Islamic terms which have been selected and then gatherd from the many English translations of the Holy Quran, Hadith and Tafseer (Interpretation) of the most prominent translators such as,Muhsin Khan, Abd-al-Hamid Siddiqui,Yusuf Ali, Pickthall, Dr. Ghali, Sahih International and Shakir, that particularly relevant to the topics of the Fasting and Pilgrimage. The significant terms and their relationships have an important role with regard to organizing the knowledge of a particular domain [26]. The researcher collected These Islamic terms and their synonyms and had approved by experts. Consequently, the question expansion technique will address a number of problems that might encounter it by using these ontological resources, including:

- Modifying the weakness of user's question in terms of selecting the words, some of the users have not enough skills to type their questions efficiently, which minimize the opportunities to retrieve accurate answers to their questions [13].
- Addressing the problem of term variation between the question and the Quranic verses, which cause mismatches between them, due to the users might use terms differ from the terms that used in the Quranic verses. Moreover, the Islamic terms are different from translation to another depending on the Translator. Therefore, the terms of user's question will be improved to be compatible with the terms of Quranic verses [14].

The following table 2 shows the words in the user's question and their synonyms from the collection of Islamic terms.

4.3 Evaluation

Since the purpose of this research is to achieve the matching between the terms of the questions and Quranic verses,therefore, the evaluation will apply to three systems based information retrieval, by using Recall. The Recall and Precision are traditional metrics which is used for information retrieval systems [21]. The first system is based on keywords matching without using the question expansion technique. The second system relies on the question expansion technique using only WordNet synonyms. The third system is also based on the question expansion technique using the synonyms from WordNet and the synonyms from the collection of Islamic terms. Experts have composed the questions used in the evaluation, where these questions cover most of the Quranic verses of Fasting and Pilgrimage. Moreover, these questionswere formulated in several styles, where it was taking into consideration the vocabularies that are used in the available translations of the Holy Quran. In these questions, a variety of Islamic vocabularies were used, which some of these vocabularies are commonly used by the readers, and some are not widely used by the users, to cover all potential styles of users' questions.

Table 2: The Islamic synonyms

Word	Synonyms
Fast	fasting, saum
Iftar	drink, eat, breaking the fast
Dawn	fajr
Ransom	fidyah , expiation, compensation
Ramadan	ramadhan, ramazan
Imsak	abstinence, abstention, privation
Pilgrimage	haj, hajj, pilgrim journey, religious journey, religious expedition
Umrah	umra, minor pilgrimage, visitation to makka
Circumambulation	tawaf, circumambulating
Maqam	muqam, place, station

Abraham	ibrahim
Pursuit	sa'ay, saee, trotting

Figure 4 shows an example of the whole process of the question expansion technique through using synonyms and generating a number of questions based on these synonyms.

4.3   Results and discussion

Based on the results of these three systems, as it shown in the following table 3, it can be note that, the evaluation of the first system which achieved based on Recall is 0.62 level. This shows that there are many of questions did not get the answersbecause of the mismatch between the words of the questions on the one hand, and the words of the translation of the Holy Quran on the other hand.In the second system, it can observe that the rate of Recall has been increased which becomes at 0.85 level, and that leads to conclude that the question expansion technique by using the WordNet has presented a lot of possibilities which led to a matching between the words of user's question on the one side, and the words of the Holy Quran translation on the other side that consequently led to retrieve the answers of these questions. However, still, not all questions find the answers. Due to the mismatch between the terms of questions and Quranic verses, which are mostly being Islamic terms.This is because the variation of the Islamic terms in the different translations and as well as, the various Islamic terms used by users such as "Ramadan", where some users and translations use "Ramadhan" or "Ramazan". In the third system, all of the terms of the questions that arepertaining to the topics of Fasting and Pilgrimage have been matched with Quranic terms, which means, all these questions retrieved answers.The use of WordNet to expand the question has a significant advantage to removing differences and mismatch between the used terms in question and the terms of the Quran, but there are many of Islamic terms do not exist in the WordNet database. Therefore, the collection of Islamic terms pertaining to fasting and pilgrimage topics will remove the differences between terms and provide a matching for each term used by the user and the existing translations of the Quran, and thus retrieve all the required verses. The process of question reformulation into equivalent questions increases the recall of the information retrieval system [21].

Table 3: Systems evaluation

System No.	Technique	Recourses	Recall
System 1	Keywords matching	-----------	0.62
System 2	Question expansion	WordNet	0.85
System 3	Question expansion	WordNet + Collection of Islamic terms	1

5.   CONCLUSION

Due to the different terms that have been used in the various Quran translations, particularly, the Islamic terms that depends on the style of the translator. As well as, the user does not know which terms used inside each translation to type its question in order to retrieve a specific answer. Therefore, this research addressed the problem of mismatch between the terms of user's question and the terms used in the translations of the Quran through using question expansion technique depending on the suggested synonyms from the WordNet and the suggested words from the collection of Islamic terms.As compared with the search systems that do not use the Islamic synonyms, it proved that the expanding of the question using the only WordNet is not enough to achieve the matching between terms. Thus, the collection of Islamic synonyms has a great significance with regard to providing a definite matching between the Islamic terms and retrieving answers to all the tested questions.
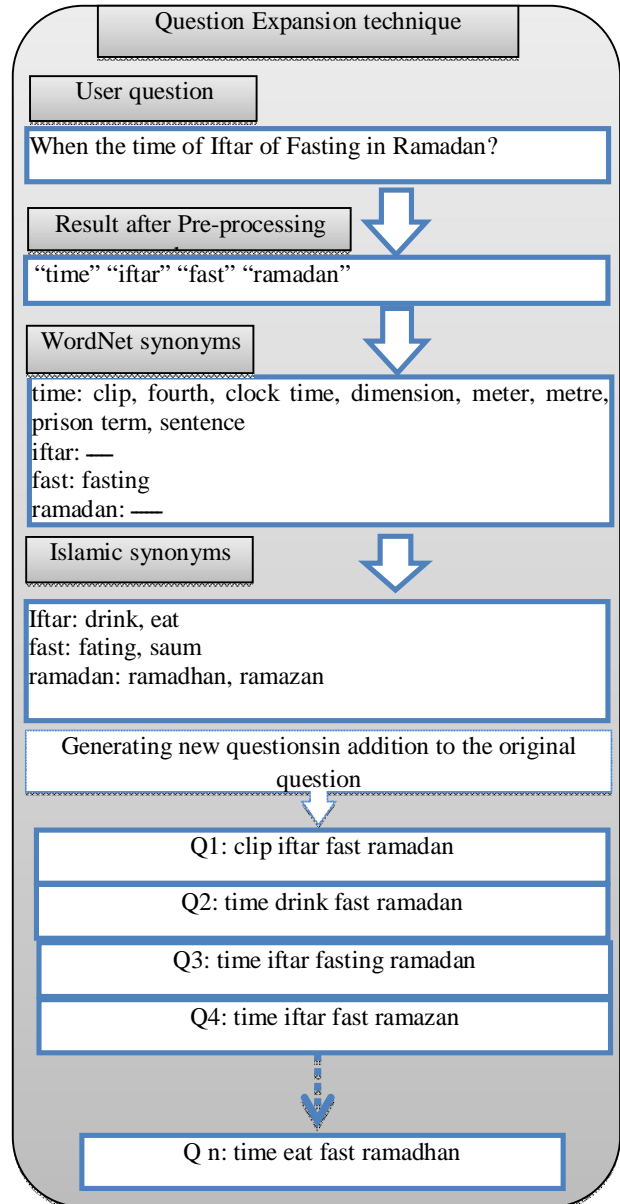

Figure 4: Example of the generated questions

REFERENCES

1.   *Komiya, K., et al., Question answering system using Q & A site corpus Query expansion and answer candidate evaluation. SpringerPlus, 2013. 2(1): p. 1-11.*

2.   Kolomiyets, O. and M.-F. Moens, *A survey on question answering technology from an information retrieval perspective.* Information Sciences, 2011. **181**(24): p. 5412-5434.

3.   Aziz, S.K.H.a.M.J.A., *Classification of Holy Quran Translation Using Neural Network Technique.* Journal of Engineering and Applied Sciences, 2018. **13**(12): p. 4468 - 4475.

4.   Tabrizi, A.A. and R. Mahmud. *Issues of coherence analysis on English translations of Quran.* in *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on.* 2013. IEEE.

5.   Aldahesh, A.Y., *(Un) Translatability of the Qur'an: A Theoretical Perspective.* International Journal of Linguistics, 2014. **6**(6): p. 23.

6.   Rezvani, R. and P. Nouraey, *A Comparative Study of Shifts in English Translations of The Quran: A Case Study on "Yusuf" Chapter.* Khazar Journal of Humanities & Social Sciences, 2014. **17**(1).

7.   Hamed, S.K. and M.J. Ab Aziz, *A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification.* Journal of Computer Science, 2016. **12**(3): p. 169-177.

8.   Alqahtani, M. and E. Atwell, *A Review of Semantic Search Methods to Retrieve Information from the Qur'an Corpus.* 2015.

9.   Shoaib, M., et al. *Relational WordNet model for semantic search in Holy Quran.* in *Emerging Technologies, 2009. ICET 2009. International Conference on.* 2009. IEEE.

10. Imran, H. and A. Sharan, *Thesaurus and query expansion.* International journal of computer science & information Technology (IJCSIT), 2009. **1**(2): p. 89-97.

11. Ishkewy, H. and H. Harb, *ISWSE: Islamic Semantic Web Search Engine.* International Journal of Computer Applications, 2015. **112**(5).

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

13. Gillet, P., et al. *Complex correspondences for query patterns rewriting.* in *Proceedings of the 8th International Conference on Ontology Matching-Volume 1111.* 2013. CEUR-WS. org.

14. Noordin, M.F., *KNOWLEDGE EXPLORATION: SELECTED WORKS ON QURAN ONTOLOGY DEVELOPMENT.* Journal of Theoretical & Applied Information Technology, 2015. **72**(3).

15. Kara, S., et al., *An ontology-based retrieval system using semantic indexing.* Information Systems, 2012. **37**(4): p. 294-305.

16. Fernández, M., et al., *Semantically enhanced information retrieval: An ontology-based approach.* Web Semantics: Science, Services and Agents on the World Wide Web, 2011. **9**(4): p. 434-452.

17. Wang, H., et al., *Conceptual representing of documents and query expansion based on ontology*, in *Web Information Systems and Mining.* 2012, Springer. p. 489-496.

18. Raj, P., *Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System.* arXiv preprint arXiv:1311.3175, 2013.

19. Hu, D., et al. *ACQA_onto: an ontology approach for restrain domain question answering system.* in *Information Science and Control Engineering 2012 (ICISCE 2012), IET International Conference on.* 2012. IET.

20. Karyawati, A.E., et al., *Ontology-based Why-Question Analysis Using Lexico-Syntactic Patterns.* International Journal of Electrical and Computer Engineering, 2015. **5**(2): p. 318.

21. Allam, A.M.N. and M.H. Haggag, *The question answering systems: A survey.* International Journal of Research and Reviews in Information Sciences (IJRRIS), 2012. **2**(3).

22. Khan, H.U., et al., *Ontology based semantic search in Holy Quran.* International Journal of Future Computer and Communication, 2013. **2**(6): p. 570.

23. Hammo, B., A. Sleit, and M. El-Haj, *Effectiveness of query expansion in searching the Holy Quran.* 2007.

24. Ramasubramanian, C. and R. Ramya, *Effective pre-processing activities in text mining using improved porter's stemming algorithm.* International Journal of Advanced Research in Computer and Communication Engineering, 2013. **2**(12): p. 2278-1021.

25. Ahmad, O., et al., *A survey of searching and information extraction on a classical text using ontology-based semantics modeling: A Case of Quran.* Life Science Journal, 2013. **10**(4): p. 1370-1377.

26. Mukhtar, T., H. Afzal, and A. Majeed. *Vocabulary of Quranic Concepts: A semi-automatically created terminology of Holy Quran.* in *Multitopic Conference (INMIC), 2012 15th International.* 2012. IEEE.