Journal of Theoretical and Applied Information Technology
10th January 2016. Vol.83. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DOMAIN-SPECIFIC ONTOLOGY-BASED APPROACH FOR ARABIC QUESTION ANSWERING

[1] MUSTEFA SHEKER, [2] SAIDAH SAAD, [3] REHAB ABOOD, [4] MOHANAAD SHAKIR

[1,2,3]Center for Artificial Intelligence Technology (CAIT), National University of Malaysia
[4]Alburaimi University Collage, Oman
E-mail: [1]alnaser.mustafa@gmail.com, [2]saidah@ukm.edu.my [3]rehab_hasan_ab@yahoo.com
[4]alnaseere1976@gmail.com

ABSTRACT

Automatic Question Answering (QA) is the process of identifying exact answer in response of question asked by the user. Several QA approaches have been proposed for many languages. One of such languages is Arabic which spoken by nearly 200 million people. Approximately 90% of Arab people are Muslims. Many Muslims need religious consultations toward specific issues. Such consultations have to be acquired from a legitimate source such as Islamic scholars. The response toward these consolations is called a Fatwa, which is considered as an Islamic point of view toward such issues. Therefore, there is a vital demand to present a QA system that has the ability to answer Muslims questions based on a collection of Fatwas. Several research efforts have been implemented to provide QA for Islamic domain. However, such efforts have used simple lists of lexicons as a source of knowledge, which leads to several limitations in terms of the correctness of the answers. This is due to the process of analyzing question would not be guided with a rich source of knowledge. Hence, this study aims to propose an ontology-based question answering approach for the domain of Islamic Fatwa. The ontology has been constructed using a collection of Fatwas, which has been collected from Ibn Uthaymeen-Prayer Fatwas. Several pre-processing tasks have been applied in order to eliminate the irrelevant data (e.g. numbers, non-Arabic letters and punctuation). Furthermore, Term Frequency-Inverse Document Frequency (TFIDF) has been used in order to provide the main concepts of the domain for ontology construction. The proposed ontology-based QA has been evaluated by comparing it with QA approach that uses simple semantic lexicon as a source of knowledge. The proposed method has outperformed the other approach by obtaining an approximately 90% of F-measure. This demonstrates that the use of ontology as a source of knowledge for QA would enrich the process of extracting the answers.

Keywords: *Question Answering, Query Expansion, Ontology, Similarity Measures, Arabic*

1. INTRODUCTION

Question Answering (QA) is the process of identifying an accurate and automatic answer that is generated by computer in response to a random question asked by the user (restricted). Information Retrieval (IR) is the process of retrieving relevant document in response of the user query [1]. It is associated with question answering in terms of retrieving the correct answer (relevant document) in response to the user's question (query) [2]. The key difference between IR and QA is that in IR user get relevant document however, he has to search in such document in order to obtain an answer to his question. Despite IR, Information Extraction is the process of identifying predefined templates from an unstructured text for instance, extracting the named entities form a given text is considered as an IE where named entities represent such templates [3]. IE has been used to capture a specific portion of documents; such portions could be used for reporting such as providing a report of daily stocks. Thus, IE may be viewed as a limited form of question answering in which the questions (templates) are static and the data from which the questions are to be answered are an arbitrarily large dynamic collection of texts.

Many challenges have faced QA such as retrieving an answer from unstructured data where the classic QA answering systems were rely on a database and the process of retrieving the answer was consider as a query task. In addition, proposing open-domain and restricted-domain QA was one of the common challenges. Basically, there are multiple differences between restricted domain

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

question answering and open domain question answering in terms of several aspects. For instance, in terms of the size of data, open domain question answering systems require huge amount of text corpus [4]. Furthermore, there are differences between open domain question answering and restricted domain question answering systems in terms of the context of domain [5]. In fact, the sense of words in restricted domain would be less than in open domain for instance, the word 'compound' has a single sense in the domain of biomedical which refers to the chemical compounds. While in the open domain, the word 'compound' may refer to chemical, linguistic (e.g. noun compound) or it may refer to group of buildings. Therefore, restricted-domain question answering systems require more specialized user, semantic and relations among entities.

On the other hand, language is considered as a significant challenge in terms of question answering where each language has its own morphology and semantic. Unlike Latin languages where Question Answering (QA) systems are relatively advanced [6-9] Question answering systems for Arabic language are very few. To our knowledge, there are few research works on Arabic Question Answering [10-12]. Mainly, it is due to the lack of accessibility to linguistic resources, such as data sets and basic NLP tools (tokenizers, morphological analyzers, etc.). In comparison with other languages, not many are the NLP tools and resources in general which are available for Arabic. Moreover, the Arabic language has a very complex morphology (inflectional and derivational characteristics) and texts suffer from the scarcity of vowels as well as the absence of capitalization. These specificities of the Arabic language introduce many processing problems related to the word tokenization, the identification and recognition and classification of named entities, etc.

Islamic Fatwa is an answer that has been provided by respected Islamic scholars in response to a question asked by Muslim individuals. Muslims have many queries regarding daily issues such as work, education, family, praying, etc. Such queries usually turn into questions that would be asked to an Islamic scholar (i.e. Sheikh) who has the knowledge about the Islamic principles. There are several resources and books for Fatwas have been published as a collection of answers for many issues. Hence, it would be interesting to propose an automatic QA system for Islamic Fatwas that has the ability to provide an answer toward an arbitrary question asked by user. Such system would be valuable for Muslim community regarding by providing accurate answer from respected Islamic scholars in response to their questions.

In fact, several QA systems have been proposed for Islamic Fatwa. However, such systems are using open-domain ontology such as Arabic WordNet. The problem of such ontology lies on the variety of definitions of concepts which are usually general (i.e. are not related to Islamic domain). Therefore, this study aims to propose a domain-specific ontology for automatic QA system of Islamic Fatwas. The proposed system has been evaluated by using two types of ontology; open-domain ontology (Arabic WordNet) and domain-specific ontology that has been constructed for this purpose.

2. RELATED WORK

One of the earliest Arabic question answering systems is the one that has been implemented by Hammo et al. [12] where the focus was on identifying the proper names in order to extract the answer. In fact, the authors have applied the Arabic morphological analyzer AraMorph in order to tokenize and normalize the text by removing the irrelevant data such as stop-words. In addition, AraMorph has been used also to identifying the tag of words such as adjective, verb, noun and others in order to identify the proposer names. Furthermore, some keywords have been used in order to indicate the answers. Similarly, Abuleil & Evens [13]. Such efforts has concentrated on identifying proper names in order to extract answer. This is regarding to the fact that the valuable information lies besides the proper names. In fact, the author has designed a rule-based approach that rely on keywords such as 'manager', 'doctor' and 'professor' and some special verbs such as 'said' and 'announced'. Such hypothesis implies the co-occurrence between these keywords and the proper names. Obviously, both efforts did not consider analyzing the question.

Abouenour et al. [10] have proposed a three-level approach for improving the passage retrieval regarding Arabic question answering using TREC questions. The authors have concentrated on the fact that a good question answering system represented by providing an accurate method for passage retrieval. In fact, the first level of the approach aims to analyze the question based on keywords and semantic correspondences. This has been performed by applying statistical methods in order to use the keywords to retrieve words that co-occurred with it. In other word, identifying the highest density passages in terms of the considered keywords. Second level is the structure-based level

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which aims to analyze the question based on the structure or the word order. For this manner, Java Information Retrieval System (JIRS) has been used as a passage retrieval module. As a result, the authors have reported an enhancement regarding to the passage retrieval.

Trigui et al. [14] have proposed a new question answering method for Arabic language which aims to answer non-complex questions or questions that required short answers using CLEFF corpus. The proposed method consists of four phases; first analyzing the question which aims to remove the stop-words in order to maintain the question with the important words. Second phase aims to analyze the anaphora that located in the question. The third phase aims to retrieve the related passages that probably contain the answer. Note that, if the selected passages do not contain the answer, a rule-based approach has been used in order to provide a list of inference represented by two words generated from the background collection. Such inference aims to replace the words with its semantic corresponding words in order to reach the answer. The fourth phase aims to select the accurate answer from the retrieved passages.

Abdelnasser et al. [15] have proposed an Arabic question answering system for Holly Quran where the system enables user to type question about Quranic verses and then response with an answer that associated with Tafseer Ibn Katheer corpus of text which contains the interpretation of Quranic verses. In fact, the proposed system consists of three main steps. The first step aims to preprocess the question by removing the stop-words, retrieving the roots of the words and identifying the POS tagging. For this purpose, MADA Arabic morphological analyzer tool has been used. The second step which is question analysis aims to classify the question into its class (e.g. who, where, what, etc.) where a Support Vector Machine classifier has been used in order to classify the questions. The third step is an information retrieval module that aims to extract the accurate passage that would contain the answer. For this manner, Quran ontology which has been presented by Atwell et al. (2011) was used to identify the Quranic concepts. In addition, multiple features have been used for this step such as TF-IDF and Named Entity Recognition approach.

Nicosia [16] have proposed an answer selection method for community question for both English and Arabic languages. In fact, the proposed method focused on the question that asked in public forums where several comments would be posted regarding the asked question. The proposed method aims to select the best answer from the comments. The proposed method consists of multiple features including lexical, syntactic and semantic features. The lexical features aims to measure the similarity between the question and its corresponding comments based on the lexical morphology of the words. For this manner, TF-IDF has been used. Syntactic features aims to measure the similarity grammatically, for this manner shallow parser and POS tagging have been used. Semantic features aims to measure the similarity based on the meaning, for this manner, Latent Semantic Analysis (LSA) has been used. Finally, a supervised machine learning method which is logistic regression has been used in order to classify the comments into correct and incorrect answers.

3. PROPOSED SYSTEM

The architecture of the proposed system as shown in Fig 1 consists of eight components which are represented as follows; *Dataset, User Question*, *Preprocessing*, *Question Analysis*, *Question Expansion*, *Specific-domain Ontology*, *Open-domain Ontology*, *Evaluation*. (i) Dataset phase refers to the data that has been collected and used for implementation. (ii) User Question phase aims to enable user to type a question related to Islamic Fatwas. (iii) Preprocessing phase aims to analyze the typed question by the user. Such analysis include preprocessing the question by eliminating the irrelevant data such as numbers, stop-words and punctuation. (iv) Question Analysis phase aims to identify similar question from the dataset using similarity measures including Cosine and Jaccard. (v) Question Expansion phase aims to expand the typed question by providing semantic correspondences using the open-domain ontology. (vi) Specific-domain Ontology which is the proposed knowledge source for the Islamic Fatwa domain and it has been constructed using TFIDF from the dataset. (vii) Open-domain Ontology which is an online ontology called Arabic WordNet [17]. Such ontology contains a large-scale semantics in Arabic including relations. (viii) Evaluation phase which aims to evaluate the proposed system based on the retrieval effectiveness. However, the following sub-sections illustrate these phases in further details.

3.1. Dataset

The dataset that has been used for this study consists of collection of Fatwas that have been collected from Ibn-Othaimeen Prayer Fatawas Book [18]. The dataset consists of four documents of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fatwas (questions and answers) that related to Islamic praying. Table 1 shows the details of the dataset.

Table 1. Dataset description

Document	Questions	Answers
Document 1	357	357
Document 2	288	288
Document 3	276	276
Document 4	266	266
Total	1094	1094

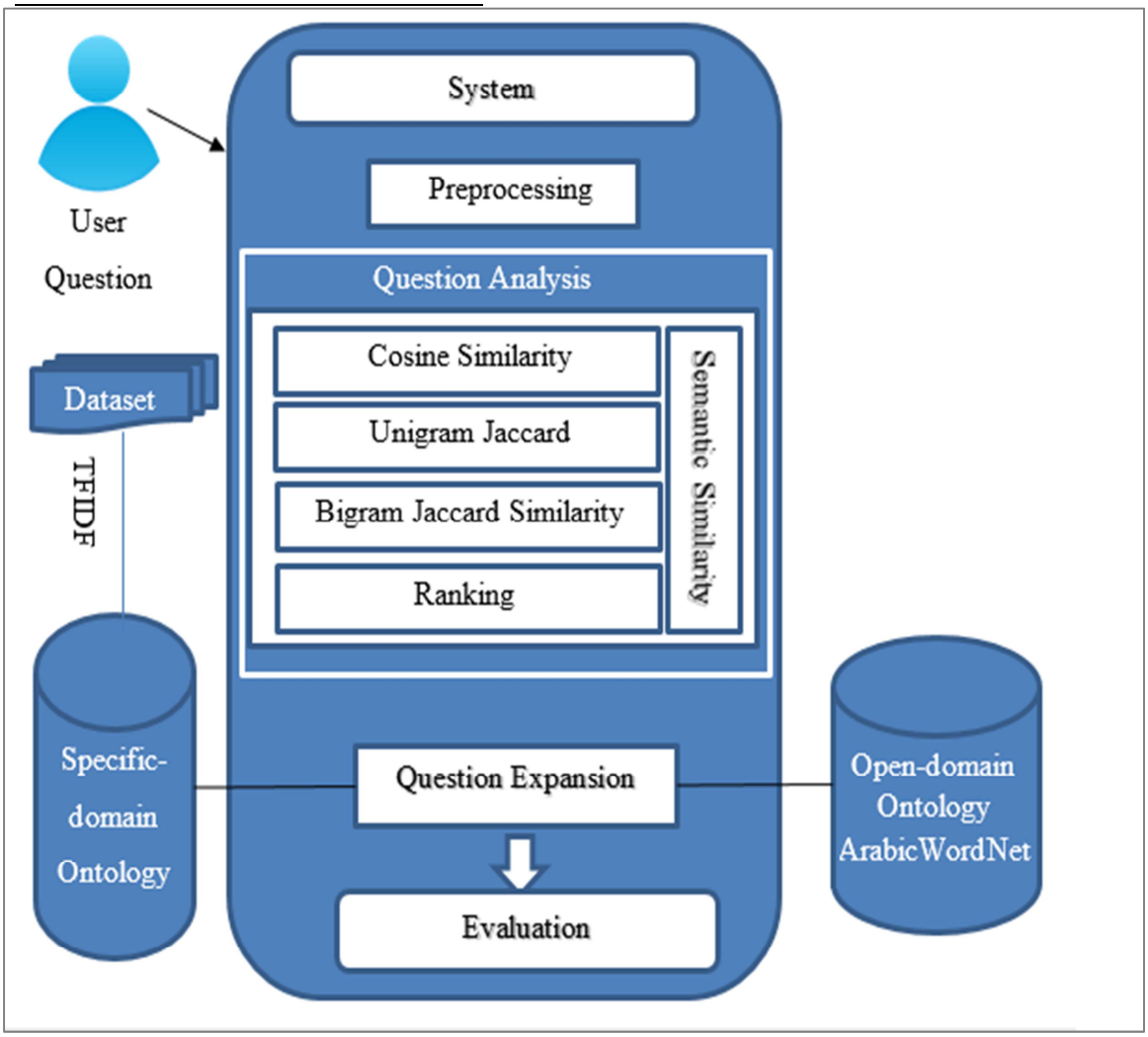


Figure1. System Architecture

Journal of Theoretical and Applied Information Technology
10th January 2016. Vol.83. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In fact, the questions and answers in the dataset are associated with each other. Therefore, it will be used in order to compare with the typed question by the user. Hence, the highest similarity between the typed questions with questions from the dataset refers that the answer will be the same. Table 2 shows a sample of question and answer that located in the dataset.

Table.2. Sample of question and answer in the dataset

Question	Translation
السؤال: هل يجوز ترك المسجد القريب والصلاة في مسجد بعيد يوم الجمعة	**Question:** Is it permissible for the worshiper to abandon the mosque which is in his area on Friday and go to another which is far away?
الجواب: يفضل أن تصلى صلاة الجمعة بالمسجد القريب مع أصدقائه وجيرانه، وذلك لخلق روح التفاعل والانسجام ولكي يشجع بعضهم الاخر. لكن إذا كان أحدهم يريد أن يصلي في مسجد بعيد لاكتساب المعرفة مثل سماع خطبة لها تأثير كبير ويزيد معرفته أكثر فلا يوجد أي اعتراض	**Answer:** It is better for him to pray with the people in his area in their mosque, so that they become acquainted with one another and in order to spread harmony between them and so that they encourage each other. But if one of them goes to another mosque in order to gain some religious benefit, such as acquiring knowledge or hearing a Khutbah which has a greater effect on him and increases his knowledge more, then there is no objection to it

3.2. User Question

This phase aims to enable the user to type a question that related to the Islamic Fatwas. Such typing process is passing through a graphical user interface. Regarding to the type of expected user, the expected user of the proposed system has to be expert in Islam which requires knowing Arabic language. On other hand, the type of question can be classified as opinionated question where the answer is based on the perspective of particular Islamic scholar.

3.3. Preprocessing

This phase aims to perform several tasks including encoding and normalization. Such tasks facilitate turning the data into a suitable form that enables further processing. Encoding aims to unify the encoding of the dataset in order to avoid character appearance problem. This problem occurs when treating Arabic letters by the default encoding (e.g. ASCI code). Therefore, UTF-8 has encoding has been used in order to convert the data into a compatible form. Whereas, normalization aims to eliminate the unwanted data such as stop-words, digits and punctuation.

3.4. Question Analysis

This phase aims to analyze the user typed question in terms of its morphology. This analysis aims to identify the most similar question from the dataset which obviously leads to acquire the exact answer. In fact, two similarity measures have been used which are Cosine and Jaccard. Cosine similarity is one of the common similarity measures, which applies to the documents in many applications such as, information retrieval, text mining and clustering [19]. To measure the similarity between documents using similarity, select two strings A and B, and used the equation of cosine similarity as below:

$$Cosine\ Similarity\ (A,B) = \frac{A.B}{\|A\| \times \|B\|}$$

The Jaccard coefficient, which is sometimes, referred as a statistical coefficient, which is used in natural language processing (NLP). Jaccard coefficient is used to measure similarity as the intersection is divided by the union of the objects [20]. For text document, Jaccard coefficient has been used to compare the sum of shared terms weight and the sum of terms weight, which are present in either of the two documents, however they must not be shared terms. The mathematical formal definition of Jaccard coefficient is expressed below:

$$Jaccard(A,B) = \frac{A.B}{|A|^2 + |B|^2 - A.B}$$

Basically, Cosine and Jaccard has been used in this study in order to measure the similarity between the user typed question with the questions from the dataset. Hence, the most similar question will be retrieved with its corresponding answer responding to the user question. The reason behind selecting these two measures lies on the different performance that could be obtained. According to Jiang et al. [21], Dice and Cosine similarity measures have a relatively same performance when dealing with short sentences. Therefore, Cosine and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Jaccard have been used as similarity measures. Note that, multiple mechanisms have been used in terms of identifying the similarity. According to Thada & Jaglan [22], Jaccard similarity reveals different performance when applying different n-gram methods (e.g. unigram, bigram, etc.). Unlike Jaccard, Cosine has shown a relatively same performance in terms of applying different n-gram methods. Therefore, Jaccard has been used carried out twice; first to measure the similarity between questions using one word (Unigram), second to measure the similarity between questions using two words (Bigram).

Obviously, applying Cosine, Jaccard for unigram and Jaccard for bigram would lead to different results of similarity among the questions. Therefore, a ranking approach has been used in order to gain the most similar question. In order to illustrate the process of ranking, suppose the user has typed a question Q_u and the question analysis phase has revealed similarities with questions from the dataset as follows:

Cosine	Q_{10}, Q_{14}, Q_{16}
Unigram-Jaccard	Q_9, Q_6, Q_{10}
Bigram-Jaccard	Q_{12}, Q_{10}, Q_6

Hence, a ranking approach has been applied that aims to calculate the frequency of candidate questions among the results. Based on the previous example, the ranking results would be as follows:

Q16	1
Q14	1
Q12	1
Q10	3
Q9	1
Q6	2

Based on the ranking results, Question no. 10 will be selected as the most similar question for the user's question, and then the answer of Question no. 10 will be brought as the exact answer of the user's question.

3.5. Question Expansion

Query expansion is the process of reformulating the query typed by the user to retrieve relevant documents [23]. Basically, the words in the query will be analyzed in terms of semantic and morphology and spilling errors. This study has applied the mechanism of expansion on the question in order to enhance the process of retrieving correct answer. In fact, two knowledge sources have been used for the expansion process. First, is the proposed specific-domain ontology, and second is an open domain-domain ontology. Such ontologies have been illustrated as follows:

3.5.1 Proposed specific-domain ontology

This study has constructed a specific-domain ontology for the field of Islamic Fatwa regarding to Islamic praying. The main classes of this ontology have been created using Term Frequency – Inverse Document Frequency (TFIDF) from the dataset. TFIDF consists of term frequency which is the number of the occurrences of the term in the document [24]. It can be calculated as follow:

$$W_d\,(t) \;=\; TD(t,d\,)$$

Where TD is the frequency of term t in document d. In addition, TFIDF contains the inverse document frequency IDF which aims to provide high weight for rare conditions and low values for common conditions (Tokunaga & Makoto 1994). The formula as follows:

$$IDF_t \;= \log n\,\left(\frac{N}{N_t}\right)$$

Where N is the number of documents in the Arabic documents and N_t is the number of documents that contain the term. Finally, TFIDF combines both the term frequency and inverse document frequency based on the following equation:

$$W_t = TF(t,d).IDF_t$$

Therefore, based on mentioned frequencies, the main classes have been created from the dataset. Hence, many sub-classes have been also created based on the main classes. Furthermore, several annotations (words related to the class) have been created in association with main classes. Table 3 depicts the details of the proposed ontology.

Table 3. The proposed ontology details

Attributes	Answers
Main classes	13
Sub-classes	134
Relations	19
Annotations	511

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

As shown in Table 3, relations refer to the correlations among the classes such as is-a, part-of and instance-of. Whereas, annotations refer to the synonyms of each class such as Human and Scholar. Such parameters have been created using experts in Arabic language.

3.5.2 Open-domain ontology

This study has used an open-domain ontology which called Arabic WordNet that has been introduced by [17]. This Ontology has been developed using a translation-based method where English WordNet lexicon has been translated into Arabic including classes, sub-classes, annotations and relations using Suggested Upper Merged Ontology (SUMO).

3.6 Evaluation

Since the evaluation will be based on retrieval effectiveness thus, the metrics that have been used is precision, recall and f-measure. Precision refers to the correctness of retrieval answers, while recall refers to the correctness of not retrieved answers. Such metrics can be calculated as the following equations:

$$Precision\ P = \frac{TP}{TP + FP}$$

$$Recall\ R = \frac{TP}{TP + FN}$$

Where TP is the number of correct retrieved answers, FP is the number of incorrect retrieved answers, and FN is the number of correct answers but not retrieved. Hence, it is possible to calculate the overall accuracy which called f-measure based on the following equation:

$$f - measure = 2 \times \frac{P * R}{P + R}$$

4. RESULTS

In fact, the evaluation has been performed using 30 questions that have been proposed by expertise in Islamic studies. Such questions are not exist in the dataset instead, similar questions in the meaning are exist in the dataset. Hence, evaluating these questions have been performed using two approaches; using Arabic WordNet and using the proposed domain-specific ontology. Table 4 and 5 depicts such results.

Table 4. Results of open-domain ontology Arabic WordNet

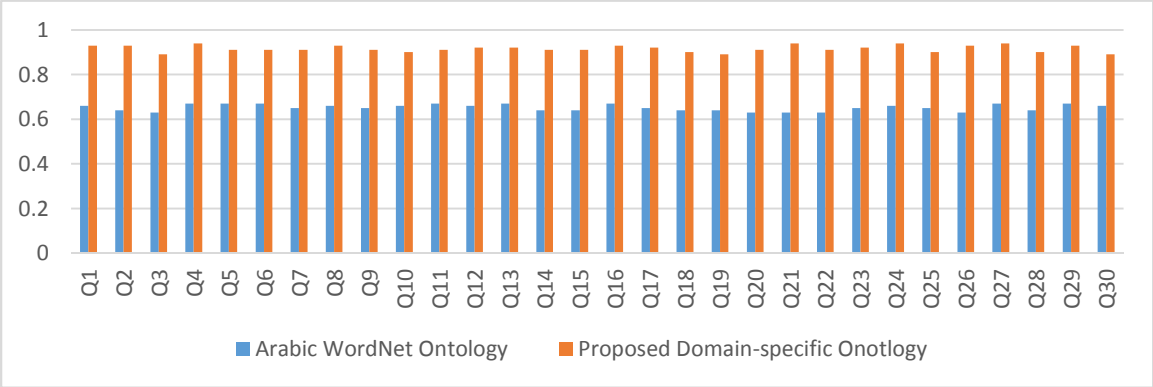
Question No.	Precision	Recall	F-score
Q1	0.73	0.60	0.66
Q2	0.72	0.58	0.64
Q3	0.70	0.57	0.63
Q4	0.74	0.61	0.67
Q5	0.74	0.61	0.67
Q6	0.74	0.61	0.67
Q7	0.72	0.59	0.65
Q8	0.73	0.60	0.66
Q9	0.73	0.59	0.65
Q10	0.73	0.60	0.66
Q11	0.75	0.61	0.67
Q12	0.74	0.60	0.66
Q13	0.74	0.61	0.67
Q14	0.71	0.58	0.64
Q15	0.71	0.58	0.64
Q16	0.74	0.61	0.67
Q17	0.73	0.59	0.65
Q18	0.72	0.59	0.64
Q19	0.71	0.58	0.64
Q20	0.71	0.57	0.63
Q21	0.70	0.57	0.63
Q22	0.70	0.57	0.63
Q23	0.72	0.59	0.65
Q24	0.73	0.60	0.66
Q25	0.73	0.59	0.65
Q26	0.70	0.57	0.63
Q27	0.74	0.61	0.67
Q28	0.71	0.58	0.64
Q29	0.74	0.61	0.67
Q30	0.74	0.60	0.66
Average	0.72	0.59	0.65

Table 5. Results of the proposed specific-domain ontology

Question No.	Precision	Recall	F-score
Q1	0.94	0.92	0.93
Q2	0.94	0.92	0.93
Q3	0.90	0.88	0.89
Q4	0.95	0.93	0.94
Q5	0.92	0.90	0.91
Q6	0.92	0.90	0.91
Q7	0.92	0.91	0.91
Q8	0.94	0.93	0.93
Q9	0.92	0.90	0.91
Q10	0.91	0.89	0.90
Q11	0.91	0.90	0.91
Q12	0.93	0.91	0.92
Q13	0.93	0.91	0.92
Q14	0.92	0.90	0.91
Q15	0.92	0.90	0.91
Q16	0.94	0.92	0.93
Q17	0.93	0.91	0.92
Q18	0.91	0.90	0.90
Q19	0.90	0.89	0.89
Q20	0.92	0.90	0.91
Q21	0.95	0.93	0.94
Q22	0.92	0.91	0.91
Q23	0.93	0.91	0.92

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Q24	0.95	0.93	0.94	Q28	0.91	0.89	0.90
Q25	0.91	0.89	0.90	Q29	0.94	0.92	0.93
Q26	0.94	0.92	0.93	Q30	0.90	0.88	0.89
Q27	0.95	0.93	0.94	Average	0.92	0.90	0.91



As shown in Table 4, the average results for all the 30 questions when using Arabic WordNet were 72%, 59% and 65% for precision, recall and f-measure respectively. Whereas, Table 5 shows that the average results for all the 30 questions when using the proposed domain-specific ontology are 92%, 90% and 91% for precision, recall and f-measure. Obviously, the proposed specific-domain ontology has outperformed the other method. This has been expected regarding to the study of [25] where the restricted-domain ontology has a significant impact on improving the effectiveness of question answering system especially when the system is so specific.

5. CONCLUSION

Islamic Fatwa has a significant role in the Muslim community where several issues have to be performed based on the Islamic rules which provided by Islamic scholars. This study proposed a question answering system for Islamic Fatwas based on a domain-specific ontology. A comparison has been held by comparing the performance of the proposed domain-specific ontology and open-domain ontology which is Arabic WordNet. The experimental results have shown that the proposed domain-specific ontology has outperformed the other approach. This demonstrates the significant impact of using domain-specific ontology for restricted-domain question answering. The future directions of this study can be represented by extending the ontology's classes to include more Islamic concepts. This has a significant impact toward containing more Fatwas. However, the future directions lies on enriching the domain-specific ontology in order to contain more concepts and terms. This can improve the process of acquiring more answers for typed questions by the user.

6. ACKNOWLEDGEMENT

This work has been supported by the University Research Grant GUP-2015-003.

REFERENCES

[1] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to information retrieval* vol. 1: Cambridge university press Cambridge, 2008.

[2] Ellen M Voorhees and Donna K Harman, *TREC: Experiment and evaluation in information retrieval* vol. 1: MIT press Cambridge, 2005.

[3] Gonçalo Simoes, Helena Galhardas, and Luısa Coheur, "Information Extraction tasks: a survey," in *Proc. of INForum*, 2009Retrieved from.

[4] Eric Brill, Jimmy J Lin, Michele Banko, Susan T Dumais, and Andrew Y Ng, "Data-Intensive Question Answering," in *TREC*, 2001, p. 90.doi:Retrieved from.

[5] Roberto Navigli, "Word sense disambiguation: A survey," *ACM Comput. Surv.*, vol. 41, pp. 1-69, 2009.doi:10.1145/1459352.1459355Retrieved from.

[6] Fangtao Li, Yang Tang, Minlie Huang, and Xiaoyan Zhu, "Answering opinion questions with random walks on graphs," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, 2009, pp. 737-745.doi:Retrieved from.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] Vanessa Lopez, Victoria Uren, Enrico Motta, and Michele Pasin, "AquaLog: An ontology-driven question answering system for organizational semantic intranets," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 5, pp. 72-105, 2007Retrieved from.

[8] Ana Cristina Mendes and Luísa Coheur, "An Approach to Answer Selection in Question-Answering Based on Semantic Relations," in *IJCAI*, 2011, pp. 1852-1857.doi:Retrieved from.

[9] Samaneh Moghaddam and Martin Ester, "AQA: aspect-based opinion question answering," in *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, 2011, pp. 89-96.doi:Retrieved from.

[10] Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso, "Three-level approach for passage retrieval in Arabic question/answering systems," in *Proc. Of the 3rd International Conference on Arabic Language Processing CITALA2009, Rabat, Morocco*, 2009Retrieved from.

[11] Yassine Benajiba and Paolo Rosso, "Arabic question answering," *Diploma of advanced studies. Technical University of Valencia, Spain,* 2007Retrieved from.

[12] Bassam Hammo, Hani Abu-Salem, and Steven Lytinen, "QARAB: A question answering system to support the Arabic language," in *Proceedings of the ACL-02 workshop on Computational approaches to semitic languages*, 2002, pp. 1-11.doi:Retrieved from.

[13] Saleem Abuleil and Martha Evens, "Extracting Names From Arabic Text for Question-Answering Systems," in *RIAO*, 2004, pp. 638-647.doi:Retrieved from.

[14] Omar Trigui, Lamia Hadrich Belguith, Paolo Rosso, Hichem Ben Amor, and Bilel Gafsaoui, "Arabic QA4MRE at CLEF 2012: Arabic Question Answering for Machine Reading Evaluation," in *CLEF (Online Working Notes/Labs/Workshop)*, 2012Retrieved from.

[15] Heba Abdelnasser, Reham Mohamed, Maha Ragab, Alaa Mohamed, Bassant Farouk, Nagwa El-Makky, and Marwan Torki, "Al-Bayan: An Arabic Question Answering System for the Holy Quran," *ANLP 2014,* p. 57, 2014Retrieved from.

[16] Massimo Nicosia, Simone Filice, Alberto Barrón-Cedeno, Iman Saleh, Wei Gao Hamdy Mubarak, Preslav Nakov, Giovanni Da San Martino, Alessandro Moschitti, and Kareem Darwish, "QCRI: Answer Selection for Community Question Answering–Experiments for Arabic and English," in *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval*, 2015Retrieved from.

[17] William Black, Sabri Elkateb, Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, Adam Pease, and Christiane Fellbaum, "Introducing the Arabic wordnet project," in *Proceedings of the Third International WordNet Conference*, 2006, pp. 295-300.doi:Retrieved from.

[18] Abdul Aziz Ibn Baz, Muhammad Al-Uthaymeen, and Rabee Al-Madkhalee, "Three Essays on the Obligation of Veiling. edited by Translated by Abu Maryam Ismaeel Alarcon Toronto: Al," ed: Ibaanah Book Publishing, 2003.

[19] Sung-Hyuk Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *City,* vol. 1, p. 1, 2007Retrieved from.

[20] Anna Huang, "Similarity measures for text document clustering," in *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008), Christchurch, New Zealand*, 2008, pp. 49-56.doi:Retrieved from http://scholar.google.com.au/scholar.bib?q=info:enBKVjSSXjQJ:scholar.google.com/&output=citation&hl=en&as_sdt=2000&ct=citation&cd=0.

[21] Yu Jiang, Guoliang Li, Jianhua Feng, and Wen-Syan Li, "String similarity joins: An experimental evaluation," *Proceedings of the VLDB Endowment,* vol. 7, pp. 625-636, 2014Retrieved from.

[22] Vikas Thada and Vivek Jaglan, "Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm," *International Journal of Innovations in Engineering and Technology,* 2013Retrieved from.

[23] Ellen M Voorhees, "Query expansion using lexical-semantic relations," in *SIGIR'94*, 1994, pp. 61-69.doi:Retrieved from.

[24] Mikio Yamamoto and Kenneth W Church, "Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus," *Computational Linguistics,* vol. 27, pp. 1-30, 2001Retrieved from.

[25] Diego Mollá and José Luis Vicedo, "Special Section on Restricted-Domain Question Answering," *Computational Linguistics,* vol. 33, 2006Retrieved from.