Question Answering Systems on Holy Quran: A Review of Existing Frameworks, Approaches, Algorithms and Research Issues

F S Utomo1,* , N Suryana2 and M S Azmi3

1 Department of Information System, Faculty of Computer Science, Universitas AMIKOM Purwokerto, Purwokerto, Indonesia
1,2,3 Center for Advanced Computing Technology (C-ACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

* Corresponding author, e-mail: fandy_setyo_utomo@amikompurwokerto.ac.id

Abstract. Question Answering System has the ability to present an answer based on a question submitted by the user in natural languages. This system consists of question processing, document retrieval, and answer extraction component. Challenge to optimize Question Answering’s system is to increase the performance of all components in the framework. The performance of all component which has not been optimized has caused to the lack of accurate answer from the systems. Based on this issue, the purpose of this research is to investigate the research gaps in the current state of existing Question Answering Systems on Holy Quran. The result of this study reveals potential research issues, namely morphology analysis, question classification, search techniques, and ontology resources.

1. Introduction

Based on data from ThoughtCo (https://www.thoughtco.com), the Muslim population in the world reached 1.8 billion people in 2017. The commandment of words of Allah stored in the holy book of Muslims which name is Quran. Furthermore, this sacred book also keeps the instruction and guidance to humankind. Human has trouble to understand the Quran content. It is because the content of the Quran has interpretation and also this holy book rich semantics of words.

Information retrieval method could be applied to holy Quran for search the information which contains in it. There are several approaches IR models could be used to provide information on the Quran, i.e. String Matching, Metadata Search, Vector Space Model, Probabilistic Model, Natural Language Processing, Latent Semantic Indexing, Text Mining, and Semantic Ontology Techniques [1].

On traditional information retrieval engines such as Google and Yahoo, they retrieve the information based on the keywords inputted by the user. These engines often provide inaccurate information to users. To handle this problem, Question Answering systems (QAS) can provide an interface where the users could express their need in natural language form. Then, this system will present accurate information based on questions inputted by the users [2]–[4]. QAS could provide specific answers, not a list of documents like in traditional search engines [5]. Confirmed by [6], [7] QAS users have the minimum number of reading since this system able to provide a direct answer based on the question posed by the users.

Challenge to optimize Question Answering Systems is to increase the performance of all modules in the framework. The performance of all component which has not been optimized has
caused to the lack of accurate answer from the systems. Based on this issue, the purpose of this research is to investigate the research gaps in the current state of existing Question Answering Systems on Holy Quran.

There are four sections below is structured as follows. Section 2 describes materials and methods. Section 3 shows the recent studies on question answering systems for Holy Quran. Section 4 presents the open research issues. Finally, Section 5 explains the research conclusion.

2. Materials and Methods
The research questions (RQ) were specified to keep the review focused. Table 1 shows the research questions and motivation discussed by this literature review.

| ID | Research Question                                                                 | Motivation                                                                 |
|----|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| RQ1| What are the primary modules inside Question Answering Systems for Holy Quran?    | To investigate the primary modules inside Question Answering Systems for Holy Quran. |
| RQ2| What kind of tools able to perform question answering systems development         | To identify the tools which able to perform question answering systems development |
| RQ3| How do the existing systems store knowledge of Quran and Tafseer?                 | To identify the existing systems storing knowledge of Quran and Tafseer.      |
| RQ4| How do the existing approaches perform Question Answering Systems Development on Holy Quran? | To identify the existing approaches perform Question Answering Systems development on Holy Quran. |
| RQ5| How to evaluate the existing Question Answering Systems on Holy Quran?            | To identify the evaluation techniques for the existing Question Answering Systems on Holy Quran. |

There are several activities which involved in the searching process for literature, i.e. digital library selection, search string determination, search string execution, and receive a list of essential literature from digital libraries. Digital libraries are used in the searching process, namely ACM digital library, Emerald, IEEE Xplore, ScienceDirect, Springer Link, Scopus, and Google Scholar.

To select the essential studies, we used several inclusion and exclusion criteria. These criteria are shown in Table 2.

| Inclusion Criteria                              | Studies discuss Question Answering systems using Quran data sets. |
|------------------------------------------------|------------------------------------------------------------------|
|                                                 | Studies discuss developing, modeling, or comparing performance of Question Answering Systems on Holy Quran. |
|                                                 | The journal version will be selected if the studies have both the proceeding and journal versions. |
| Exclusion Criteria                              | Studies not written in English.                                    |

3. Recent Studies on Question Answering Systems for Holy Quran
This Sub-Section is organized as follows. Sub-Section 3.1 describes significant article publications between January 2014 and March 2017. Sub-Section 3.2 presents the research findings from previous studies. Sub-Section 3.3 shows the question answering question essential components. Sub-Section 3.4 shows tools to perform question answering systems development. Sub-Section 3.5 presents the existing systems store knowledge of Quran and Tafseer. Sub-Section 3.6 presents the current
approaches to perform Question Answering Systems development on Holy Quran. Finally, Sub-Section 3.7 shows evaluation techniques to test the question answering systems.

3.1. Significant Article Publications
Based on the searching results, there is nine literature which relevant to the search condition. In 2014 there was three number of studies, in 2015 there was two number of studies, while in January 2016 until March 2017 there was four number of studies. Figure 1 shows the number of literature on each digital libraries.

![Figure 1. Search Results on Digital Libraries](image)

According to Figure 1, we collected literature from specific sources, i.e., Conference and Journal. From the mentioned Figure, we could conclude that research for question answering systems on Holy Quran was rarely investigated.

3.2. Research Findings from Previous Studies
Previous studies have developed QAS on Quran with several languages and several types of question input characteristic. Table 3 describes some previous studies in QAS on Holy Quran.

| Study | Language | Question Input | Domain        |
|-------|----------|----------------|---------------|
| [8]   | Arabic   | Factoid        | Restricted    |
| [9]   | Indonesian | Factoid       | Restricted    |
| [10]  | Arabic   | Simple statement | Restricted    |
| [11]  | Arabic   | Factoid        | Restricted    |
| [12]  | English  | Simple statement | Restricted    |
| [13]  | English  | Simple statement | Restricted    |
| [14]  | Indonesian | Factoid (who, where, when) | Restricted    |
| [15]  | Arabic   | Factoid        | Restricted    |
| [16]  | Arabic   | Simple statement | Restricted    |

There are two types of question input in QAS, i.e., factoid (what, where, who, when, which, and how much/many questions) and non-factoid (how and why question) [17]. Based on Table 3, most of the previous studies focus on factoid as a question input, while Arabic is the most language used in previous works. All previous research focuses on a restricted domain which means the questions and answers are coming from particular topics inside Holy Quran.
Each research results in previous studies have the main findings. These findings covered about a new framework for question answering systems on Holy Quran, algorithms, techniques, ontologies, and databases. Several previous studies have main findings in a new framework for question answering systems. Research by [8] developed Al-Bayan an Arabic question answering system for the Holy Quran. [9] also developed a new framework for question answering systems on Indonesian Quranic translation. They developed a new framework by adding rule-based scoring for each question type (who, when, and where). Then, research by [10] developed an advanced search technique using semantic modeling. A study by [11] developed a new technique called Arabic Ontology Extractor (AOE) for ontology extraction from Arabic text to store in the ontology. They also developed Arabic Quranic Ontology (AQO) which consisted of 380 concepts and 50 relations. [12] expanded the relationship (object property and data property) which is linked the concepts in QAC ontology. Research by [13] had two main findings. First, they developed a collection of Islamic terms for synonyms sets, which has been collected from many English translations of the Holy Quran, Hadith, and Tafsir that related to the themes of Fasting and Pilgrimage. Second, they developed a question answering system on English Quran translation based on Neural Network for verses classification. A study by [14] developed a semantic-based Question Answering System for Indonesian Translation of Quran. Research by [15] had two main findings. First, they developed Arabic Quranic ontology (http://quranontology.com/) which containing 234 concepts of living creations, 219 concepts of events, and 69 concepts of places mentioned in the Quran. Second, they developed the semantic-based Question Answering System for Arabic Quran. The last one, a study by [16] also had two main findings. First, they developed Quran Database which containing Arabic Quranic, 8 English translations of Quran, 4 Tafsir, Quran word dictionary, revelation reasons, concepts in the Quran, Named Entities based on Quran domain. Second, they developed an Arabic Quranic Semantic Search Tool based on ontology (AQSST). This semantic search tool consists of several primary components, i.e., an ontology of Quranic, database of Quranic, Natural Language Analyser, semantic search engine, a search engine based on words matching, and Scoring and Ranking Model.

3.3. Question Answering System Essential Components

Based on previous studies, question answering systems on Holy Quran consist of several modules, i.e., question analysis/processing, document retrieval/processing, and answer extraction module. Figure 2 describes the QAS essential components.

![Figure 2. QAS Essential Components](image)

Based on Figure 2, question processing component has a role in translating the natural language query into a form that could be processed by the document retrieval module. Document retrieval component is a module which provides a technique to identify candidate documents that hold the relevant answer to the query. Finally, the answer extraction component receives the set of passages from the document retrieval component, then determine the best answers for the user.
3.4. Tools to Perform Question Answering System Development

To develop QAS on Holy Quran, most previous researchers have used several tools. A study by [8] used MADA (Morphological Analysis and Disambiguation for Arabic) which has been developed by [18]. This tool used for Arabic text pre-processing, such as POS tagging, lemmatization, disambiguation, stemming and glossing. Furthermore, they also used Apache Lucene [19] and LingPipe. Next, studies conducted by [10], [12] used Apache JENA for building Semantic Web and Linked Data applications, and Protege to develop an ontology. Besides using Apache JENA and Protege, a study by [15] also used Stanford NLP Segmenter for token pre-processing, and Arabic Toolkit Service for Arabic stemmer. Besides using Protege, research conducted by [11] also used Stanford Parser for POS tagging. Subsequently, a study by [13] used WEKA for Named Entity Recognition (NER) in verses classification. Summary about tools utilization on each researcher has described in Table 4. All tools which are used in the previous study, are available online and free to use.

### Table 4. Tools Utilization Summary

| Study | Tools               | Function                                                                 |
|-------|---------------------|--------------------------------------------------------------------------|
| [8]   | MADA                | Arabic pre-processing toolkits                                           |
|       | Apache Lucene       | Full text searching, spellchecking, hit highlighting, tokenization, ranked searching, pluggable ranking models including the Vector Space Model. |
|       | LingPipe            | Named Entity Recognition (NER)                                           |
| [10]  | Protege             | Tools to develop ontology                                                |
|       | Apache JENA         | Java framework for building Semantic Web and Linked Data applications    |
| [11]  | Protege             | Tools to develop ontology                                                |
|       | Stanford Parser     | POS tagging                                                              |
| [12]  | Apache JENA         | Java framework for building Semantic Web and Linked Data applications    |
|       | Protege             | Tools to develop ontology                                                |
| [13]  | WEKA                | Named Entity Recognition                                                 |
| [15]  | Protege             | Tools to develop ontology                                                |
|       | Stanford NLP Segmenter | Token pre-processing                                                  |
|       | Arabic Toolkit Service | Arabic Stemmer                                                        |
|       | Apache JENA         | Java framework for building Semantic Web and Linked Data applications    |

As shown in Table 4, previous works used tools to support question processing module to analyze and transform natural language query into a format which could be processed on the next component, document processing module. Moreover, previous research also used tools such as Protege to develop knowledge representation, then used Apache JENA to retrieve information from the knowledge. This retrieval process is performed on the document processing module.

3.5. Datasets used for Question Answering Systems

In question answering system development, some researchers could use existing Quran databases. However, some researchers developed a new Quran database. Research conducted by [8], [10], [12], [14] used existing Quran databases. While, studies by [9], [11], [13] created new Quran database. However, studies by [15], [16] used existing and created new Quran database.

Quran Vocabulary (QVOC) ontology research resulted from [20], and Quranic Arabic Corpus (QAC) ontology research resulted from [21], these ontologies are most widely used by researchers in question answering systems for Holy Quran. Research by [15] used QVOC ontology to cover words morphology. Then, a study by [12], [14] used QAC ontology. QVOC and QAC ontologies were used...
by [10] in their research. Besides using both ontologies, a study by [16] also used ontologies from [22], and Quran annotated with Pronominal Anaphora (QurAna) from [23]. Ontology from [22] was also used by [8]. Besides using this ontology, they also used ontology from [24]. Furthermore, research by [13] used Surah Al-Baqarah from English translation of the Quran by Abdullah Yusuf Ali. A study by [9] also used Surah Al-Baqarah from Indonesian translation of the Quran as a data sets. Quran database from research results [9] was used by [14] as a data sets. And the last one, study by [11] used AlMaany dictionary as a data sets for synonym sets. Summary about data sets which are used on previous studies is described in Table 5.

**Table 5. Data Sets in Previous Studies**

| Study | Data Sets                                                                 | Quran Storage | Tafseer Storage | Relations Between Quran and Tafseer |
|-------|---------------------------------------------------------------------------|---------------|-----------------|-------------------------------------|
| [8]   | 1. Ontologies: [22], [24]. 2. Ibn-Kathir and Al-Jaza-iri Tafseer books.   | Ontology      | Ontology        | They used one ontology which containing Qur'an and Tafseer. Under each concept in their ontology, they saved the related verses with their Tafseer. |
| [9]   | Al-Baqarah from Indonesian translation of the Quran                       | Relational database | -              | -                                   |
| [10]  | 1. QVOC and QAC ontologies. 2. Jalalayn and Muyassar tafseer.             | Ontology      | Ontology        | They used one ontology which containing Qur'an and Tafseer. Under each concept in their ontology, they saved the related verses with their Tafseer. |
| [11]  | AlMaany dictionary                                                        | Ontology      | -               | -                                   |
| [12]  | QAC ontology                                                              | Ontology      | -               | -                                   |
| [13]  | 1. Surah Al-Baqarah from English translation of the Quran by Abdullah Yusuf Ali. 2. The collection of Islamic synonyms from the many English translations, hadith, and tafseer. 3. WordNet | Relational database | Relational database | They used three relational databases for Quran, Islamic synonyms, and WordNet. |
| [14]  | 1. Relational database from [9]. 2. Develop Indonesian translation of Quran (ITQ) ontology from QAC ontology. 3. Tafseer of Indonesian translation of Quranic verses. | Ontology      | Relational database | They used the ITQ ontology and relational database from [9]. |
| [15]  | QVOC and Quran ontology                                                   | Ontology      | -               | -                                   |
| [16]  | 1. They aligned the four main ontologies, i.e. QVOC, QAC, QurAna, and ontology from [22]. 2. Quran DB: Arabic Quranic, 8 english translations of Quran, 4 tafseer, Quran word dictionary, revelation reasons, concepts in Quran, Named Entities based on Quran domain. | Ontology      | Relational database | Ontology used in Semantic search model (SSM), while Relational DB used in Keyword Search Model (KSM). |

Based on Table 5, in [14] research, they didn't use Tafseer of Indonesian translation of Quranic verses because after the pre-processing task had been done, they found many noises came along and caused some problems. This condition led them to withdraw Tafseer database from their systems.
3.6. Approaches Used for Question Answering Systems Development on Holy Quran

Generally, there are five approaches to perform question analysis, document processing, and answer extraction inside the QAS, i.e., linguistic, statistical, semantic, rule-based pattern matching, and hybrid method [17]. Question processing, document retrieval, and answer extraction technique in previous works are described in Table 6.

### Table 6. Question Analysis, Document Processing, and Answer Extraction Technique of Previous Studies

| Study | Question Processing | Document Processing | Answer Extraction |
|-------|---------------------|---------------------|-------------------|
| [8]   | 1. Question pre-processing with the linguistic approach.  
2. Question classification with Support Vector Machine (SVM). | 1. Semantic Interpreter using machine learning as in [25] that maps fragments of text into a weighted vector.  
2. Cosine similarity to select the top scoring verses. | Answer extraction with the linguistic and statistical approach. |
| [9]   | Question pre-processing with the rule-based and linguistic method. | Keyword-matching technique and word match scoring function is applied to count number of similar words between question and document. | 1. Relevant documents are getting processed by rule-based scoring component to get final score.  
2. Find the correct answer within the highest scored document. |
| [10]  | Question pre-processing with the linguistic and semantic method. | SPARQL query |  |
| [11]  | Question pre-processing with the linguistic model. | 1. Keywords expansion  
2. SPARQL query creation  
3. SPARQL query execution |  |
| [12]  | Question pre-processing with the linguistic model. | 1. Identify and select noun concepts with a linguistic model.  
2. Noun expansion using WordNet.  
3. SPARQL query creation  
4. SPARQL query execution |  |
| [13]  | Question pre-processing with the linguistic and semantic model | Artificial Neural Network to classify the verses of Al-Baqarah and generate relevant verses from the Holy Quran | 1. Extract the answers using N-gram technique.  
2. Words Matching scoring function to determine the best answer. |
| [14]  | Question pre-processing with the linguistic and statistical model | Document processing with the semantic and statistical method. | Answer extraction with the linguistic and statistical approach. |
| [15]  | Question pre-processing with the linguistic and semantic model | 1. Entity Mapper using N-gram comparison  
2. Entity Disambiguation using ontology  
3. SPARQL query creation and execution |  |
| [16]  | Question pre-processing with the linguistic and semantic model | Document processing with the semantic and/or words matching technique | Answer extraction with the statistical model |

3.7. Evaluation Techniques to Test the Question Answering Systems

Some previous studies have been tested using specific measurement scales. Tests were conducted toward Question Classification/named entity recognition/Expected Answer Type accuracy and overall systems. They used Precision, Recall, F-measure, and F-score to test the components and whole systems. To test the question classification, [8] used Precision, Recall, and F-measure. Then to tested QAS accuracy, they used TopN accuracy from [26]. A study by [9] used ten questions for each question type (who, when, where) to test the output accuracy. Meanwhile, research by [10] examined their output accuracy based on a comparison between output and actual data. To evaluate their system
performance, [11] and [15] used Precision and Recall. Besides using Precision, a study [13] also used F-score measurement.

4. Open Research Issues
Based on the recent development of QAS on Holy Quran, there are several open issues which could be highlighted in it. These issues are languages, question classification/Expected Answer Type/Named Entity Recognition technique, keywords matching technique, and ontology resources.

A. Morphological Analysis
Natural Language Processing (NLP) technique is applied at the pre-processing stage on question analysis processing. NLP is used to parse the text and to perform morphology analysis, such as sentences splitter, tokenizer, syntactic information provision or Part of Speech (POS) tagger, and to deduct a noun phrase (NP chunker). Every language has a different written form, grammar, vocabulary, and syntax [27–29]. According to this condition, the NLP technique for particular language has a method to perform a morphological analysis which is different from other languages.

B. Question Classification
Supervised learning algorithm could be used for question classification. However, this learning algorithm demands a big training corpus in order to classify the data, so classifier has a high performance. Drawbacks from supervised learning, the classifier accuracy would weaken if the training data size is small. [30, 31]. Challenge in question classification is what technique could be used in a small dataset for high classification accuracy.

C. Keywords Matching Technique
String Matching, Metadata, Vector Space Model, Probabilistic Model, Natural Language Processing, and Text Mining technique are the type of Full-Text Searching [32]. The weakness of Full-Text Searching isn't able to obtain information based on the input word synonyms because it depends on words matching the search query with the word in the data source. This technique isn't able to produce accurate information because it can't represent the complete semantic query [33-35] such as synonym, homonym, polysemy, term, and concepts. This situation has led to unforeseen consequences such as the result of information search is irrelevant.

D. Ontology Resources
Ontology resources is a factor will affect the QAS accuracy level to provide relevant answers to the users. In question answering systems on Holy Quran, this collection could be Tafseer, Hadith, and revelation reasons. If ontology only contains Quran textual, then question answering systems would provide the answers from Quran. This method is called by the literal or textual approach. However, if ontology consists of Quran and other resources like Tafseer and Hadith, then question answering systems would provide the answers from this knowledge. This method is called by the contextual approach. The contextual approach interprets the Quran by observing the semantic or contextual information related to the Quran, such as the information obtained from the Hadith and Tafseer. Quran ontology should keep the contextual meaning of the Quran verses so the QAS could provide better information and insight to the users.

5. Conclusion
In this research, the primary modules inside QAS on Holy Quran, an existing approach to store the knowledge of Quran and Tafseer, an existing method to conduct QAS development for Holy Quran, and the evaluation techniques to test the current QAS on Holy Quran have been discussed. Many research opportunities are still available along this line and further investigations for morphological analysis in a different language, question classification, search techniques, and ontology resources.

Acknowledgments
The authors would like to thank the financial support from Universitas AMIKOM Purwokerto; and Center for Advanced Computing Technology (C-ACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM) for their assistance in this research.
References

[1] J. Euzenat, and P. Shvaiko, “Ontology matching”, 2nd Edition. Springer Berlin Heidelberg, 2013.
[2] A. Sayed, and A. Al Muqrishi, “An efficient and scalable Arabic semantic search engine based on a domain specific ontology and question answering,” International Journal of Web Information Systems, vol. 12, no. 2, pp. 242–262, 2012.
[3] A. C. Mendes, L. Coheur, J. Silva, and H. Rodrigues, “JUST.ASK — A MULTI-PRONGED APPROACH TO QUESTION ANSWERING,” International Journal on Artificial Intelligence Tools, vol. 22, no. 1, pp. 1–34, 2013.
[4] S. Saqaeeian, M. Shakibafakhri, and M. Roshanzadeh, “Proposed an Optimal Search Algorithm to Find the Best Answer in a Question Answering Systems,” Analele Universitatii 'Eftimie Murgu', vol. 21, no. 1, pp. 311–320, 2014.
[5] J. Bian, Y. Liu, E. Agichtein, and H. Zha, “Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media,” in Proceedings of the 17th international conference on World Wide Web, 2008, pp. 467–476.
[6] D. Hristovski, D. Dinevski, A. Kastrin, and T. C. Rindflesch, “Biomedical question answering using semantic relations,” BMC Bioinformatics, vol. 16, p. 6, 2015.
[7] Jovita, Linda, A. Hartawan, and D. Suhartono, “Using Vector Space Model in Question Answering System,” in Procedia Computer Science, 2015, vol. 59, pp. 305–311.
[8] H. Abdelnasser et al., “Al-Bayan : An Arabic Question Answering System for the Holy Quran,” in 9th International Workshop on Semantic Evaluation, 2014, pp. 57–64.
[9] R. H. Gusmita, Y. Durachman, S. Harun, A. F. Firmansyah, H. T. Sukmana, and A. Suhaimei, “A rule-based question answering system on relevant documents of Indonesian Quran Translation,” in International Conference on Cyber and IT Service Management (CITSM), 2014, pp. 104–107.
[10] A. Hakkoum and S. Raghay, “Advanced Search in the Qur'an using Semantic modeling,” in IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), 2015, pp. 1–4.
[11] H. Shmeisani, S. Tartir, A. Al-Na’ssaan, and M. Naji, “Semantically Answering Questions from the Holy Quran,” in International Conference on Islamic Applications in Computer Science And Technology, 2014, pp. 1–8.
[12] A. R. Yauri, R.A. Kadir, A. Azman, M. Azrifah, and A. Murad, “Applying Ontological Engineering Approach for Arabic Quran Corpus: A comprehensive Survey,” in IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), 2015, pp. 620–627.
[13] S. K. Hamed and M. J. A. Aziz, “A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification,” Journal of Computer Science, vol. 12, no. 3, pp. 169–177, 2016.
[14] S. J. Putra, R. H. Gusmita, K. Hulliyah, and H. T. Sukmana, “A semantic-based Question Answering System for Indonesian Translation of Quran,” in Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services, iiWAS ’16, 2016, pp. 504–507.
[15] A. Hakkoum and S. Raghay, “Semantic Q&A System on the Qur’an,” Arabian Journal for Science and Engineering, vol. 41, no. 12, pp. 5205–5214, 2016.
[16] M. Alqahtani and E. Atwell, “Arabic Quranic Search Tool Based on Ontology,” in 21st International Conference on Applications of Natural Language to Information Systems (NLDB), 2016, vol. 9612, pp. 478–485.
[17] F. S. Utomo, N. Suryana, and M. S. Azmi, “Question Answering System: A Review on Question Analysis, Document Processing, and Answer Extraction Techniques,” Journal of Theoretical and Applied Information Technology, vol. 95, no. 14, pp. 3158-3174, 2017.
[18] N. Habash, O. Rambow, and R. Roth, “MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization,”
in *Proceedings of the Second International Conference on Arabic Language Resources and Tools*, 2009, pp. 102–109.

[19] U. Schindler, B. Brauer, and M. Diepenbrock, “Data Information Service based on Open Archives Initiative Protocols and Apache Lucene,” in *German e-Science Conference*, 2007, pp. 1–11.

[20] M. A. Sherif and A. C. Ngonga Ngomo, “Semantic Quran,” *Semantic Web*, vol. 6, no. 4, pp. 339–345, 2015.

[21] K. Dukes, “Statistical Parsing by Machine Learning from a Classical Arabic Treebank,” University of Leeds, Leeds, 2013.

[22] N. Abbas, “Quran search for a Concept Tool and Website,” The University of Leeds, 2009.

[23] M. Sharaf and E. S. Atwell, “QurAna: Corpus of the Quran annotated with Pronominal Anaphora,” in *8th International Conference on Language Resources and Evaluation (LREC’12)*, 2012, pp. 130–137.

[24] E. Atwell, C. Brierley, K. Dukes, M. Sawalha, and A. Sharaf, “An Artificial Intelligence approach to Arabic and Islamic content on the internet,” in *Proc NITS’2011 National Information Technology Symposium, King Saud University, Saudi Arabia*, 2011, no. January, pp. 1–13.

[25] E. Gabrilovich and S. Markovitch, “Computing semantic relatedness using Wikipedia-based explicit semantic analysis,” in *Proceedings of the 20th international joint conference on Artificial intelligence*, 2007, pp. 1606–1611.

[26] C. D. Manning, P. Raghavan, and H. Schutze, “An Introduction to Information Retrieval,” Online, no. c, p. 569, 2009.

[27] B. Andi-pallawa, “A Comparative Analysis between English and Indonesian Phonological Systems,” *International Journal of English Language Education*, vol. 1, no. 3, pp. 103–129, 2013.

[28] B. R. Chiswick and P. W. Miller, “Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages,” *Journal of Multilingual and Multicultural Development*, vol. 26, no. August, pp. 1–18, 2005.

[29] A. U. Rahayu, “Differences on Language Structure between English and Indonesian,” *International Journal of Languages, Literature and Linguistics*, vol. 1, no. 4, pp. 257–260, 2015.

[30] A. H. Asiaee, P. Doshi, T. Minning, and R. L. Tarleton, “A framework for ontology-based question answering with application to parasite immunology,” *Journal of biomedical semantics*, vol. 6, p. 31, 2015.

[31] D. Tomás and J. L. Vicedo, “Minimally supervised question classification on fine-grained taxonomies,” *Knowledge and Information Systems*, vol. 36, no. 2, pp. 303–334, 2013.

[32] J. Beall, “The Weaknesses of Full-Text Searching,” *Journal of Academic Librarianship*, vol. 34, no. 5, pp. 438–444, 2008.

[33] X. Li, J. Martinez, and G. Rubio, “A New Fuzzy Ontology Development Methodology (FODM) Proposal,” *IEEE Access*, vol. 4, pp. 7111–7124, 2016.

[34] J. Zhong et al., “Conceptual Graph Matching for Semantic Search,” *Lecture Notes in Computer Science*, vol. 2393, pp. 92–106, 2002.

[35] K. Saravanakumar and M. Moturi, “Semantic based Personalized Framework for Information Retrieval,” *International Journal of Computer Applications*, vol. 20, no. 4, pp. 14–17, 2011.