AL-QURAN ONTOLOGY BASED ON KNOWLEDGE THEMES

A. Ta’a¹*, Q. A. Abed², and M. Ahmad¹

¹School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, Kedah, Malaysia
²Al-Furat Al-Awsat Technical University, Iraq

Published online: 17 October 2017

ABSTRACT

Islamic knowledge is gathered through the understanding the Al-Quran. It requires ontology which can capture the knowledge and present it in a machine readable structured. However, current ontology approaches is irrelevant and inaccuracy in producing true concepts of Al-Quran knowledge, because it used traditional methods that only define the concepts of knowledge without connecting to a related theme of knowledge. The themes of knowledge are important to provide true meaning and explanation of Al-Quran knowledge classification. The main aims of this paper are to demonstrate the development of ontology Al-Quran and method used for searching the Al-Quran knowledge using the semantic search approach. Expert review has been applied to validate the ontology model and evaluate the relevance and precision of searching results.

Keywords: Al-Quran ontology; knowledge retrieval; knowledge representation; semantic web; RDF/OWL

Author Correspondence, e-mail: azman@uum.edu.my
doi: http://dx.doi.org/10.4314/jfas.v9i5s.57
1. INTRODUCTION

Al-Quran is the eternal miracle of Islam. It is the complete and best guide to living one's life and seeking Allah's pleasure. Al-Quran consisted of 30 divisions (Juz), 114 chapters (surah), 6236 verses (ayah), and less than 80,000 words. The verse may contain one or more sentences to form the verse. A group of verses will form a chapter with the given names (e.g., Al-Baqarah, Ali-Imran, An-Nisa). The concept of Al-Quran knowledge is defined by using the Quran hierarchy as determined by the sequence of divisions, chapters, and verses. These are the most important components for presenting the verse as final references of the particular Al-Quran concepts in a systematic manner. Understanding the Al-Quran required appropriate knowledge and teachers, which is sometime difficult to be available at all times. However, by using computer technology, the learners can learn and understand the Al-Quran easily, especially through web and mobile-based applications. Mostly, the approaches to presents the Al-Quran knowledge is using tree-structure hierarchy, which is limited to identifying the actual meaning of the concept in the Al-Quran [18].

This research used an ontology approach to represent Al-Quran concepts that can be classified and organized according to a specific theme. Islamic scholars have described the Al-Quran as the holy book of Muslims and this book teaches moral, purification, good deeds, and as well as those forbidden by the Almighty Allah. The Al-Quran provides guidance to mankind, promotes justice between one another, and provides guidance on how to live on earth [2] [22]. A related study described the Al-Quran as a source of information on any subject matter concerning the world and the hereafter. Therefore, knowledge in the Al-Quran cannot be compared with scientific books because the former provides real and deep discussions of matters under examination. A related study described that the Al-Quran as a source of knowledge on any subject matter concerning the world and the hereafter [18]. Allah said in the following verses:

هَذَا قُرْآنٌ إِنَّهُ مُبَيَّنٌ وَهُدًىٰ وَرَحْمَةٰ لِّلْمُتَّقِينَ

This [Qur'an] is enlightenment for mankind and guidance and mercy for a people who are certain [in faith]

(Al-Quran 45:20)
Currently, the search and retrieval of knowledge in the Al-Quran is either irrelevant or inaccurate [5]. The searching method is not based on knowledge classification, and this aspect becomes a major shortcoming of existing research issues that are related to the knowledge retrieval in the Al-Quran ontology. Ontology uses classes, properties, and individual structure to represent the semantics of Al-Quran knowledge [16]. The semantic or concept is referred to the hierarchy relationship of division (juz), chapter (surah), and verse (ayah) in the Al-Quran. Normally, the reference to the verses in Al-Quran is defined by division, chapter, and verse, which is the most important step in applying an ontology approach to perform a semantic search method [3] [4]. The semantic search method requires a well-classification of verses of Al-Quran in information technology [29]. Several studies indicated that well-classification of knowledge was defined from the background knowledge domain [32] [33]. Therefore, further studies are required to produce a semantic searching method for Al-Quran ontology that aims to classify and develop the Al-Quran ontology based on themes, and used to develop the semantic searching method accordingly.

2. LITERATURE REVIEW
Understanding the Al-Quran knowledge is a major challenge for Computer Science and Artificial Intelligence research area [7]. Knowledge representation is a systematic approach to learn and understand a science based on a number of classification methods. Some methods of classification adopted by researchers such as taxonomies, hierarchical, tree structure, paradigm and constrained facet analysis in the delegate, discover and construct new knowledge [15]. Currently, the researchers using the ontology to represent knowledge in a wider context, specifically in the domain of Al-Quran. It is important to present the Al-Quran knowledge in a systematic manner because the Quran knowledge is categorized in terms: “muhkamat” (i.e., semantic sentence is clearly understood) and “mustasyabihat” (i.e., semantic sentence is not clearly understood) that require a high level of guided explanations by the Islamic scholars.

An effort to create information systems that enable Al-Quran knowledge to be accessed online has been implemented by many researchers and system developers. Among them [6] has built the Quran ontology based on information system in common Arab language and corpus. The
ontological structure is focused on the relationship of concept in Al-Quran as defined in Arabic. [28] has developed ontology Al-Quran for Malay language and be source for determining the related concept of Malay terms in Al-Quran, and [29] has developed knowledge of Islam based on the Al-Quran by using ontology. These studies still infancy and much more work needs to be done to complete the ontology of Al-Quran. The application for the Al-Quran ontology still being developed based on traditional methods (i.e., using entity-relationship modeling). Moreover, most Al-Quran ontology is developed based on the English and Arabic languages, and therefore, not a complete application can allow searching and learning with the semantic based method to be implemented.

Ontology is defined as “an explicit formal specification of common concepts” and a branch of philosophy that involves the study of the types and structures of objects, properties, events, and processes of relations in every area of reality [31]. Descriptions of concepts and relationships can exist as an agent or a community of agents. This definition is consistent with the use of ontology as a set of concept definitions. Studies applied ontology to provide solutions in various fields, both in the social and natural sciences. Moreover, many types of ontology editors (e.g., Protégé-OWL) can be used to construct ontology. Thus, the efficient use of ontology editors during development is important while performing ontology visualization [18] [29]. Researchers have argued that many types of editors are fit for the construction of an ontology system such as Protégé-OWL because of their robustness and flexibility of use [5]. An earlier study by [12] emphasized that the use of Protégé-OWL editor during ontology is acceptable while retrieving and searching for an exact word in a pool of texts. However, the appropriate selection and efficient use of editors are important while conducting ontology visualization and development, and thus, should be given consideration. Hence, the Protégé-OWL editor should be considered in the development and construction of an ontology-based system and in the context of its flexibility.

In general, the reasons for developing ontology are to share a common understanding of information among people or software agents, to reuse of domain knowledge, to make domain assumptions explicit, to separate domain knowledge from the operational knowledge, and to analyse the domain knowledge [3]. Moreover, the research in ontology in Islamic field is not

new in the epistemology of information technology, due to its ability to describe Al-Quran verses in a common and meaningful ways [2]. Previous researchers have used ontology for prayer (Solat) classification [29], and using the Wordnet model for Al-Quran semantic search [18].

3. INFORMATION RETRIEVAL

Query formulations by users through keywords or spatial filters are helpful in presenting standard-based catalogs [10]. Therefore, the metadata fields that fit into the query depend on the metadata schema and the query functionality of the service used for accessing the metadata. Researchers have indicated that the two types of semantic heterogeneity, namely naming heterogeneity (synonyms) and cognitive heterogeneity (homonyms), which were used to discover related words or phrases from an information pool [11]. However, free-text entries are unfit to capture the semantics of a query or item [1].

3.1 Knowledge Discovery and Retrieval

The goal of knowledge retrieval from a large volume of information is to provide user support for interpreting and obtaining adequate information, which could be achieved through an ontology-based approach. Adopting the said approach to retrieve information solves the challenges of free-text search facilities in catalogs and supports an intuitive interpretation of specific words. This approach will need to enable the navigation of differences in meaning [11]. Moreover, [32] suggested the use of explicit context models to re-interpret information for new applications. Thus, ontology gained popularity in information science because of its capacity to explicate contextual information [19] [20] [24].

3.2 Knowledge Representation

The Al-Quran is the most widely read holy books in the world. Interest in automating knowledge extraction and retrieval from religious literature has led to the development of a number of searching applications to provide the ability to retrieve knowledge by using keywords [14]. A plethora of websites with Al-Quran and Hadith texts, search tools, reference materials, maps, and others have all been designed around natural language for human readers. However, none of these websites provide a standardization, which is important for the machine to process the information. [35] has presented some of the tools and programs that
use keyword-based extraction architecture to model and retrieve data. Other web applications such as [13] has used the same keyword-based analysis for data retrieving.

3.3 Ontology-Based Information System

Ontology-based knowledge exchange is a popular issue for various research groups [21]. Ontology has been treated as a good structure for storing and sharing knowledge. Ontology is a tool for exploring, inferring, and describing web sources [27]. Nevertheless, ontology has also been accepted in other fields, such as in electronic commerce and in creating terminological services in the health sector [17]. To date, ontology has become an essential tool for computer science applications and has been extended to satisfy the new vision and next generation of the World Wide Web (WWW), that is the Semantic Web. The aim of ontology is to build data structures for semantic web technology in a way that is easily understood by human and machines [34].

In ontology-based information processing, the unstructured or semi-structured natural language text is extracted and presented the information through an ontology-guided mechanism [9]. The ontology creation may be conducted manually, automatically, or semi-automatically depending on the researcher's choice and research condition. The vision of ontology-based information processing includes a number of complementary disciplines, such as machine learning, natural language processing, data mining, and so on.

The global readership of the Al-Quran has increased attention on knowledge extraction and retrieval from religious literature and has led to the creation of knowledge retrieval by using keywords [25]. Several websites with Quranic and Hadith texts, search tools, reference materials, and so on have also been designed around natural language for human readers. Several tools and programs that use keyword-based extraction architectures to model and retrieve data have been introduced for searching the Al-Quran knowledge. [26] has developed an ontology for the Bible and was created several classes of concept in the Bible, with each class having characteristics that define the concept and examples related to the concept that explained certain classes have subclasses. For example, the class book of the Bible has a relationship with the class chapter via the abbreviation properties.

3.4 Knowledge Searching for Ontology Al-Quran

A large number of Quranic databases exist in digital form and provide root verse search [18].
The database processes the morphological analysis of query verse as input and provides the root verse as output. The searching of related words in the Al-Quran has led to the creation of a keyword-based searchable interface indexed by Surah number. This interface assists users in browsing the Al-Quran and searching with translation and commentary. A multilingual Quran Software provides Arabic and English Quranic commentaries. Different translations in French, German, Spanish, Urdu, Malay, Indonesian, Japanese, Tamil, Hausa, Turkish, and Indonesian are also available on many sites equipped with a query-word-based searching facilities. The software by Harf Information Technology provides a subjective search facility, but only in the Arabic language. This software also provides an exact match search for words, terms, parts of verses, and even some consecutive verses. Technically, this software provides the ability to search static files in a way that the verses are pre-linked to a topic or sub-topic. Thus, semantic search in the Al-Quran is sometimes based on Internet searches that reveal some works on the Al-Quran ontology [34].

4. RESEARCH METHODOLOGY

Several researchers have developed the Islamic ontology as a solution to the problems on a basis of stability, homogenous, and peaceful of social ordinary environment. They also said that the commentary of Al-Quran and Hadith are always required in order to clarify and put a right perspective of the Al-Quran ontology. Ontology has been used to define the concept in the Al-Quran and classify and cluster it into phrases, terms, or verses. Ontology also shows the relationship between classes and between individuals. Developing ontology is similar to defining a set of data and their schema in a database for application programs to manage and retrieve knowledge from it. Problem-solving methods, domain-independent applications, and software agents were used ontology and knowledge bases to build the database schema from the ontology [3] [4].

Information or knowledge retrieval is the process of gathering data resources that are relevant to associate data from a set of knowledge resources. The searching can be based on metadata, full text or content-based indexing. The proposed ontology-based approach for knowledge retrieval in the Al-Quran will be discussed detail in the following sections. The approach consists of two sub-stages: (i) development of an Al-Quran ontology, and (ii) development of
ontology searches for Al-Quran. Fig. 1 shows the research approach.

![Fig.1. The research approach](image)

### 4.1 Al-Quran Ontology Development.

The development of Al-Quran ontology in stage 1 has been implemented by using Protege-OWL based on Al-Quran ontology model as shown in Fig. 2 [14].

![Fig.2. Al-Quran ontology model](image)

All the main themes and sub-sub-themes were extracted and derived from the classification of Al-Quran's contents [23], and were endorsed by Islamic scholars from the Islamic Center, University Utara Malaysia (UUM), who served as panel for references, validations, and user reviews. The specification of ontology definitions is defined in three hierarchy levels. The first level is class definition, which is the three main classes (i.e., Allah, Angels, and Unseen) were defined based on the selected themes. The second level is properties definition, where the object and data proprieties is was originally taken from the Al-Quran and derived from the classification of Al-Quran's contents. The third level is individual definition, which is based on the classification of Al-Quran's contents, and has been discussed and endorsed by Islamic
scholars as relevant to the main themes and classes that derived from the Al-Quran. The appropriate terms used to explain and achieve the relationships between the main themes (i.e., Allah, Angels, and Unseen) with all the derived sub-themes were properly defined. Fig. 3 shows how the individuals were defined between classes and properties.

Ontology Searches for Al-Quran. This research proposed the used database system for implementing the searching mechanism of the Al-Quran ontology. The entry point of this process starts after the RDF/OWL file output from the Al-Quran ontology that produced from Protégé-OWL was translated as tables in the database, and all contents in the RDF / OWL file were saved into the database. Table 1, Table 2 and Table 3 show examples the database schemas. The database stored contains of RDF/OWL file, which included the references of verses in Al-Quran. The searching mechanism is based on actual terms, which is converted to SQL query on the database. The web-based interfaces were designed and developed by using Java Server Pages (JSP) to facilitate user queries or inputs. Users may perform a search for certain keywords and select the main theme, sub-theme and sub sub-theme that relevant to the keyword.

Fig. 3. Individuals derived between classes and properties
The searching approach is based on pattern matching of keyword with the database contents. The entry point of this approach starts after the RDF/OWL file output from the ontology developed by using Protégé-OWL was translated into a relational database (i.e., MySQL), where the RDF/OWL metadata (as shown in Table 1, Table 2 and Table 3) was saved into the database. The result of searching was listed in ascending order, and ready to be picked for presenting the detail results. Fig. 4 shows the searching flow chart, and the algorithm for the searching approach.

The prototype of searching application is developed to ensure the ontology specification and searching algorithm can be implemented through computer programming and deployment.
The prototype system is developed using Java programming language, JENA framework, Java Server Pages (JSP) and Hyper Text Markup Language (HTML). The JENA framework is used for ontology manipulation and it was selected because free and open source framework for building ontology-based application. Indeed, JENA provides an API to extract data from and write to the RDF files such as RDF/OWL file for Al-Quran ontology. Moreover, the Java programming language provides developer opportunity to create and deploy applications and services to the end users. Fig. 5 and Fig. 6 shows the example interfaces for the searching application.

6. EVALUATION AND RESULTS

The Al-Quran ontology is validated by two Islamic scholar from the Islamic Center, UUM. Both scholars were asked to check the correctness of ontology specification, which was defined according to the Al-Quran ontology model as shown in Figure 2. The complete
specification of Al-Quran ontology has been defined based on selected themes (i.e., Allah,
Angels and Unseen). The experts were asked to check and corrected the theme definition,
specific relationship between themes, and related to the relevant verses of Al-Quran. The
experts have endorsed the validity of ontology specification as shown in Table 4.

Consequently, the purpose of searching evaluation is to ensure the ability searching algorithm
as shown in Figure 4 to retrieve knowledge in Al-Quran, and can be accurately produce the
results according to the themes or sub-themes. The search method is tested by using a
pre-defined test case, and the purpose of using a test case is to ensure all test functionalities
are easily traced and be notified any changes occurred [8]. The test case prepares for this
method is basically aimed to test the accuracy of searching method. The inputs (i.e., theme,
sub-theme and sub sub-theme) for searching are predetermined and it's given to the user in the
testing exercise. Nine (9) queries have been used as examples of the test case as shown in
Table 5.

**Table 4. Ontology specification for validation**

<table>
<thead>
<tr>
<th>Themes</th>
<th>Relationship</th>
<th>Sub-theme</th>
<th>Relationship 2</th>
<th>Sub-sub-theme</th>
<th>Correct</th>
<th>No Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allah</td>
<td>has</td>
<td>characteristics</td>
<td>can be knowing</td>
<td>Universe and his creation</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angels</td>
<td>has</td>
<td>the duties</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unseen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5. Test case for searching**

<table>
<thead>
<tr>
<th>No. of query</th>
<th>Theme</th>
<th>Sub-Theme</th>
<th>Sub-sub-Theme</th>
<th>Number of Verses in RDF/OWL file</th>
<th>Keywords in search</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Allah</td>
<td>Knowing</td>
<td>The one character for mankind</td>
<td>21</td>
<td>Infinite character</td>
</tr>
<tr>
<td>2</td>
<td>Allah</td>
<td>Knowing</td>
<td>The universe and his creation</td>
<td>23</td>
<td>Universe</td>
</tr>
<tr>
<td>3</td>
<td>Allah</td>
<td>Characteristics</td>
<td>Existence</td>
<td>15</td>
<td>Existence</td>
</tr>
<tr>
<td>4</td>
<td>Angels</td>
<td>The duties</td>
<td>Organize all matters</td>
<td>12</td>
<td>Matters</td>
</tr>
</tbody>
</table>

**Table 6. Measurement for searching results**

<table>
<thead>
<tr>
<th>No</th>
<th>Theme</th>
<th>Sub-Theme</th>
<th>Sub sub-Theme</th>
<th>Keywords in search</th>
<th>No of Times for the keywords</th>
<th>No of Retrieved Verses</th>
<th>No of Retrieved Correct Verses</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Allah</td>
<td>Knowing</td>
<td>The one character for mankind</td>
<td>Infinite character</td>
<td>5</td>
<td>10</td>
<td>90.4</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Allah</td>
<td>Knowing</td>
<td>The universe and his creation</td>
<td>Universe</td>
<td>5</td>
<td>4</td>
<td>80</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Allah</td>
<td>Characteristics</td>
<td>Existence</td>
<td>Existence</td>
<td>15</td>
<td>15</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
By giving a keyword, the relevant verses that related to themes, sub-themes, and sub-sub-themes were retrieved. Then, the results produced from the search were evaluated on their precision and recall measurement. These two measurements are commonly used to measure the performance of the information retrieval [33]. For examples, based on query number 1 in Table 4, the search is “Innate character”. Thus, the calculation of recall is 90.4% and precision is 95% show that the query result for “the innate character” have a high percentage of accuracy according to the given theme, sub-theme and sub-sub-theme. The calculation is using the following formula:

\[
\text{Recall} = \frac{\text{No. of relevant records retrieved}}{\text{No. of relevant records retrieved} + \text{No. of relevant records not retrieved}} \times 100\% \\
= \frac{19}{19 + 2} \times 100\% = 90.4\%
\]

\[
\text{Precision} = \frac{\text{No. of relevant records retrieved}}{\text{No. of relevant records retrieved} + \text{No. of irrelevant records retrieved}} \times 100\% \\
= \frac{19}{19 + 1} \times 100\% = 95\%
\]

In summary, the result related to the query based on theme “Allah” was calculated on their precision and recall is shows in Table 6. The results have shown the highest percentage of both precision and recall measurements.

Precision and Recall are the basic measurement used for evaluating the search method. Based on the results from the evaluation, the high accuracy of the searching method was related to the used by the database. The database stored contents from the RDF/OWL, which is also having a major contribution in improving the searching accuracy. Therefore, higher percentage of precision and recall have described the correctness of the searching method used in the Al-Quran ontology. The percentage can be increased if more themes or concepts are used in the experiments. Consequently, the correctness of the searching results can facilitate the learners to learn Al-Quran in an efficient way (i.e., online learning).

Generally, the success of online learning can be measured by how much information can be gathered or retrieved from the online sources [30]. Much information retrieved from the
searching will provide more options to understand a particular knowledge. The relevant and correctness of the information retrieved will determine the success of the searching method. This paper also evaluated the success of searching functionality in learning for a particular theme (i.e., Judgment Day). The query for searching is performed into two different methods: traditional theme-based and ontology-based. Both methods were implemented in the same computer platform (i.e., Windows 8 operating system). For traditional theme-based approach, Microsoft SQL Server was used as a database system. Table 2 shows the benchmark of searching results from both methods.

### Table 7. Benchmark of searching functionality

<table>
<thead>
<tr>
<th>Method/Item</th>
<th>Traditional Theme-Based Method</th>
<th>Ontology-Based Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Theme</td>
<td>Faith</td>
<td>Faith</td>
</tr>
<tr>
<td>Theme(s)</td>
<td>Judgment Day</td>
<td>Judgment Day</td>
</tr>
<tr>
<td>Query Language</td>
<td>SQL (run on .Net apps)</td>
<td>SPARQL (run on TWINKLE)</td>
</tr>
<tr>
<td>Number of Records/Number of N-Triples</td>
<td>16 records</td>
<td>38 N-Triples</td>
</tr>
</tbody>
</table>

Table 7 shows the results that the ontology-based method has produced 38 triples/records compared to traditional theme-based method has produced 16 records. This benchmark has showed that the ontology-based searching methods were produced higher records/triples compared to traditional theme-based method. The matching results will give a wide range of knowledge to the users of learning the Al-Quran. Detail explanations for each of the information retrieved can be obtained from the various chapters of verses. Even though this benchmark only performed for one theme, the results explained that the semantic-based search method has produced several answers from the query. This provides rich information for users to be selected and learned. However, further research need to be done in order to exactly identify the relevant (recall) and correctness (precision) of searching functionality.

### 7. CONCLUSION

The Al-Quran is the most widely read holy books in the world. Interest in automating
knowledge extraction and retrieval from religious literature has led to the development of a number of searching applications with the ability to retrieve knowledge on the basis of keyword, sentence or theme. This research proposed theme-based approach for knowledge retrieval in Al-Quran ontology. The proposed approach consists of two sub stages: development of the Al-Quran ontology, and development of a semantic searching method. This research has achieved the development of Al-Quran ontology, and also developed a method that facilitated the search in Al-Quran ontology by using semantic approach. Based on the evaluation, the searching method has retrieved the Al-Quran knowledge with relevant contents and acceptable accuracy. Further work will develop applications for learning Al-Quran knowledge that adopting this Al-Quran ontology model.

8. ACKNOWLEDGEMENTS
This work was supported by the Research Acculturation Grant Scheme (RAGS). The authors wish to thank the Ministry of Higher Education of Malaysia (MOHE) and Research Innovation and Management Center (RIMC) of Universiti Utara Malaysia (UUM) for providing the opportunities to perform this research.

9. REFERENCES


[14] Islam Ware. Islam Ware, 2016


[23] Roche K. Classification of Al Qur’an’s contents, 2011


[26] Boisen S. Semantic bible, 2013


**How to cite this article:**