Ontology-Based Approach for Knowledge Retrieval in Al-Quran Holy Book

Azman Ta’a1, Qusay Abdullah Abed2, Bashah Mat Ali3, Muhammad Ahmad4

1,3School of Computing, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia,
2Al-Furat Al-Awsat Technical University, Iraq, qusy.
4College of Business, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia,

ABSTRACT

Knowledge of Al-Quran is gathered through the interpretation of Division (Juz), Chapter (Surah), and Verse (Ayah) from the representation of ontology structure. It requires a searching method, which can find true knowledge of Al-Quran. However, current searching approaches in Al-Quran ontology face several fundamental problems, such as irrelevant and inaccuracy for producing true concepts and verses of knowledge. Moreover, current approaches used conventional methods such as taxonomy, hierarchy, or tree structure, which only define the concepts of knowledge without linking to a relevant theme of knowledge. The themes of background knowledge are important to provide another possibility of knowledge explanation and to ensure the results from the searching were selected the true meaning of Al-Quran knowledge. Therefore, the main aims of this paper are to demonstrate the searching method for retrieving the relevant and accurate verses in Al-Quran by using ontology approach.

Keywords: Knowledge Retrieval, Knowledge Representation, Ontology, Semantic Web, RDF/OWL Framework, Holy Book, Al-Quran

I. INTRODUCTION

Islamic scholars have described the Al-Quran as the holy book of Muslims. This holy book teaches moral, purification, and good deeds, 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 and with neighbors [1,2]. A related study described the Al-Quran as a source of information on any subject matter concerning the world and the hereafter [3]. Knowledge in the Al-Quran cannot be compared with scientific books because the former provides real and deep discussions of matters under examination [1,3]. A related study described that the Al-Quran as a source of knowledge on any subject matter concerning the world and the hereafter (Shoaib et al., 2009). As Allah said in the following verses:

This [Qur'an] is enlightenment for mankind and guidance and mercy for a people who are certain [in faith].
(Al-Quran 45:20)

The search and retrieval of knowledge in the Al-Quran is either irrelevant or inaccurate [4]. Currently, the searching method is not based on knowledge classification [4]. This paper argues, this aspect is the major shortcomings of existing research on issues that related to the searching for knowledge in the Al-Quran by using ontology. Ontology uses classes, properties, and individual structure to represent the semantic of Al-Quran knowledge [5]. The semantic or concept is referred to as the hierarchy of division (juz), chapter (surah), and verse (ayah). Normally, reference to the verses in Al-Quran is defined by division and chapters, which is the most important step in applying an ontology approach to perform a semantic search method [24]. The searching for Al-Quran knowledge requires a well-classification of verses in information technology [6]. Several studies indicated that well-classification of knowledge defined from the background knowledge domain, and further studies are required to produce a semantic
searching method for Al-Quran ontology [1,2]. Therefore, the aims of this paper is to classify and develop the Al-
Quran ontology based on themes, and used to define the semantic searching method.

II. LITERATURE REVIEW

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 [7]. 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 [3,6].

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 [30]. An earlier study by [4] 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 [3,6]. Hence, the Protégé-OWL editor
should be considered in the development and construction of an ontology-based system and in the context of this
research due to 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 analyze the domain knowledge [24]. Moreover, the research for
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 approach for prayer
classification (Solat) [6], using the Wordnet model for Al-Quran semantic search [29].

III. INFORMATION RETRIEVAL

3.1 Knowledge Discovery

Query formulations by users through keywords or spatial filters are helpful in presenting standard-based catalogs
[8]. 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), are used to
discover related words or phrases from an information pool [9]. However, free-text entries are unfit to capture the
semantics of a query or item [10].

3.2 Knowledge 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, Sminia and Stuckenschmidt [12] 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 [13-15].

3.3 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 [25]. A plethora of websites with Quranic 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 any standardization, which is important in the machine processing of
information. [26] present some of the tools and programs that use keyword-based extraction architecture to model
and retrieve data. Numerous other web applications, such as those in [27] and [25], use the same keyword-based
analysis.
3.4 Ontology-Based Information System

Knowledge exchange is a popular issue for various research groups [16]. Thus, ontology has been treated as a good structure for storing and sharing knowledge. Ontology is built for tools for exploring, inferring, and describing web sources [17]. Nevertheless, ontology has also been accepted in other fields, such as in electronic commerce and in creating terminological services in the health sector [17, 18]. To date, ontology has become an essential tool for computer science applications. It has also been extended to satisfy the new vision and next generation of the World Wide Web, that is, the Semantic Web. The aim of ontology is to build data on the Semantic Web in a way that is easily understood by machines [19]. Ontology-based information processing unstructured or semi-structured natural language text through an ontology-guided mechanism to extract certain types of information and present the output [20]. Ontology creation may be conducted manually, automatically, or semi-automatically depending on the researcher's choice and research condition. The vision of ontology learning 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 research attention on knowledge extraction and retrieval from religious literature and has led to the creation of knowledge retrieval by using keywords [21]. 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 in searching the Al-Quran. Boisen [22] has developed an ontology from the Bible and was created several classes for concept in the Bible, with each class having characteristics that define the concept and examples related to the concept. Certain classes have subclasses. For example, the class book of the Bible has a relationship with the class chapter via the abbreviation properties.

3.5 Knowledge Searching for Ontology Al-Quran

A large number of Quranic databases exist in digital form and provide root verse search [3]. 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 Tafsir. 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 facility. The software by Harf 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[19].

III. RESEARCH METHODOLOGY

[23] has built the Islamic Ontology as a solution to all problems and the basis of a stable, homogenous, and peaceful social ordinary. They also said that the Hadith and Tafsir were the ways to acquire clarification and perspective toward 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 [24]. 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: (1) development of an Al-Quran ontology, and (2) development of ontology searches for Al-Quran. Figure 1 shows the main architecture for the proposed approach.

![Figure 1. The Research Approach](image)
4.1 Al-Quran Ontology Development

The first development stage has been implemented by using Protege-OWL. All the main themes and sub-sub-themes were extracted and derived from the classification of Al-Qur'an's contents (Original Format) [28], and were endorsed by Al-Quran experts from the Islamic Center in the University Utara Malaysia (UUM), who served as panel for references, validations, and user insights. 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 is the object and data proprieties was originally taken from the Al-Quran and derived from the classification of Al-Qur'an's contents (Original Format) [28]. The third level is individual definition, which is based on the classification of Al-Qur'an's contents, and has been discussed and endorsed by Al-Quran experts 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. Figure 2 shows how the individuals were defined between classes and properties.

4.2 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.

Table 1. Reference for Theme

<table>
<thead>
<tr>
<th>Theme</th>
<th>Position in Al-Quran</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOnteness</td>
<td>Surah (21): Ayat (21)</td>
</tr>
<tr>
<td>AllahKnowsAllThatIsInTheHeart</td>
<td>Surah (16): Ayat (23)</td>
</tr>
<tr>
<td>OrganizeAllMatters</td>
<td>Surah (77): Ayat (4)</td>
</tr>
<tr>
<td>OrganizeAllMatters</td>
<td>Surah (77): Ayat (1)</td>
</tr>
<tr>
<td>AllahKnowsTheKeyOfAllThatIsHidden</td>
<td>Surah (6): Ayat (59)</td>
</tr>
</tbody>
</table>

Figure 2. Individuals derived between classes and properties

Table 2. Reference for Chapter

<table>
<thead>
<tr>
<th>Chapter Name</th>
<th>Number of Chapter in Al-Quran</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Fatihah</td>
<td>1</td>
</tr>
<tr>
<td>Al-A'raf</td>
<td>21</td>
</tr>
<tr>
<td>Al-An'am</td>
<td>3</td>
</tr>
<tr>
<td>Al-Nisa'</td>
<td>4</td>
</tr>
<tr>
<td>Al-Kahf</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3. Reference for Verse

<table>
<thead>
<tr>
<th>Text</th>
<th>Short Text</th>
<th>Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>The inanimate character for mankind</td>
<td>TheInanimateCharacterForMankind</td>
<td>Inanimate character</td>
</tr>
<tr>
<td>Existence</td>
<td>Existence</td>
<td>Existence</td>
</tr>
<tr>
<td>Glorify Allah and prostrate before him</td>
<td>GlorifyAllahAndProstrateBeforeHim</td>
<td>Prostration</td>
</tr>
</tbody>
</table>
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 search will be listed in ascending order, and ready to be picked for presenting the detail results. Figure 3 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. Figure 5 shows the example interfaces for the searching application.

IV. EVALUATION AND RESULT

The Al-Quran ontology has been validated by two Islamic scholar from the Islamic Center, UUM. Both experts were asked to check the correctness of ontology specification which was defined using Protégé-OWL. The complete specification of Al-Quran ontology based on selected themes (i.e., Allah, Angels and Unseen) were validated by the experts. The experts were asked to check and corrected the theme, the specific relationship between themes, and relate to the relevant verses of Al-Quran. The experts have endorsed the ontology specification as shown in Table 4.
Table 4. Ontology Specification for Validation

<table>
<thead>
<tr>
<th>Themes</th>
<th>Relationship 1</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 characteristics</td>
<td></td>
<td>by way</td>
<td>the innate character for mankind.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>by way</td>
<td>the universe and his creation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angels</td>
<td>has the duties</td>
<td></td>
<td>is</td>
<td>organize all matters.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is</td>
<td>the hearts of the Prophets and the Faithful.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is</td>
<td>praying for the Faithful and ask for forgiveness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is</td>
<td>glorify Allah and prostrate before him.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is</td>
<td>recording the actions of mankind.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is</td>
<td>blowing of the Trumpet.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is</td>
<td>conveying the revelation to the prophets.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is</td>
<td>guarding the hell and punish its occupants.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unseen</td>
<td>is that</td>
<td>Allah knows the keys of unseen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is that</td>
<td>Allah knows the affairs of the hidden.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is that</td>
<td>Allah knows all that is in the heart.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The purpose of searching evaluation is to ensure the ability to retrieve knowledge in Al-Quran can be accurately achieved according to the themes or sub-themes. The search is tested by using a pre-defined test case, and the purpose of using a test case is to test all functionalities and to easily trace any changes occur [31]. The test case that is prepared for this method is basically aimed to test the accuracy of searching method. The inputs (theme, sub-theme and sub sub-theme) for searching are predetermined and it’s given to user in the testing exercise. Nine (9) queries have been used as examples of the test case. Table 5 shows part of the prepared test case.

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 innate character for mankind.</td>
<td>21</td>
<td>Innate 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>Creation</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>

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 [32]. For examples, based on query number 1 in Table 4, the search is “Innate character”. Thus, the calculation of precision and recall is shown in Figure 7. The results for 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.
Ontology-Based Approach For Knowledge Retrieval In Al-Quran ...

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\%

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\%

Figure 6. Calculation for Precision and Recall

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

<table>
<thead>
<tr>
<th>No. of query</th>
<th>Theme</th>
<th>Sub-Theme</th>
<th>Sub-sub-Theme</th>
<th>Keywords in Search</th>
<th>No. of Variants for the Keyword</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 innate character for mankind</td>
<td>Innate character</td>
<td>21</td>
<td>20</td>
<td>19</td>
<td>90.4</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>Allah</td>
<td>Knowing</td>
<td>The universe and his creation</td>
<td>Universe</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>66.6</td>
<td>80</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>13</td>
<td>86.6</td>
<td>86.6</td>
</tr>
</tbody>
</table>

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.

V. CONCLUSION

The Al-Quran is the most widely read 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 keywords. This research proposed an ontology-based approach for knowledge retrieval in Al-Quran. The proposed approach consists of two sub stages: development of the Al-Quran ontology, and development of a searching method. Finally, this research have achieve to develop Al-Quran ontology, develop a method that facilitated the search in Al-Quran ontology, and retrieve Al-Quran knowledge with high accuracy.

VI. 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.

REFERENCES

 Ontology-Based Approach For Knowledge Retrieval In Al-Quran ...


