International Journal of Engineering & Technology, 7 (4.15) (2018) 126-129

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Retrieval Performance for USIM's Quranic Search Engine

Nurlida Basir[1,2*], N. F. Nabila[1], Nurzi Juana Mohd Zaizi[1], Madihah Mohd Saudi[1,2], Farida Ridzuan[1,2], Sakinah Ali Pitchay[1,2]

[1]Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.
[2]CyberSecurity and Systems Research Unit, Islamic Science Institute (ISI), Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.
**Corresponding author E-mail: nurlida@usim.edu.my*

Abstract

The Quran covers every field of life namely medical, political, social, economic, and others. The extraction of quranic knowledge is a challenging task, as the Quran is rich in its linguistic and multi-layered meanings, difficult if done without the use of other resources such as the Hadith and Tafsir by Muslim Scholars. Several Quranic ontologies have been developed, however, most of these only focus on certain domains and language, rather than the content of the Quran. In many cases, the searching cannot retrieve the relevant fields. Discovering the fields of a keyword searched in the Quran can become a guidance to enhance knowledge in a particular domain. In this paper, we propose that the keyword searched in the Quranic ontology can be applied for extracting the relevant fields. For this purpose, exploratory search has been done. The sample domain ontology based on the medical and economic fields mentioned in the Quran has been developed in TopBraid Composer. Queries have been run to depict the proper role of ontology. Then, certain recommendation for the proposed Quranic ontology attaining semantic search from all domains and the result by searching from verse to domain, or domain to verse, has been proposed.

Keywords: *Information Retrieval; Ontology; Quran; Semantic.*

1. Introduction

The Quran and the Hadith contain information about many domains and include all kinds of knowledge, for example, knowledge in science, which is not yet known and explored by man. It is a golden opportunity for us, especially Muslims to explore and extract the mystery of science in the Quran in order to lead in the discovery and exploration of science in the future [1]. Figure 1 depicts the separation of Naqli and Aqli knowledge in education in the form of "Dualism". This would not be beneficial for the Ummah (people), especially Muslim. This cause imbalance and shortcomings in the education [2].

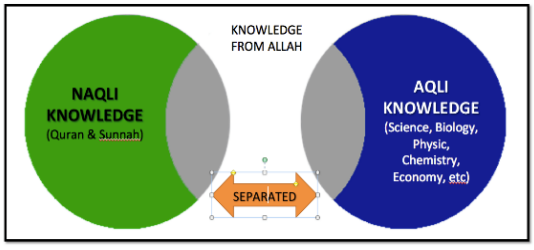


Fig. 1: Naqli and Aqli Knowledge

Therefore, this work is trying to reduce the gaps by developing an index Naqli-Aqli through quranic ontology that will allow experts in the field of aqli to make a discovery of knowledge from the Quran and Hadith for community well-being purposes.
Recently, several Quranic ontologies have been developed. However, most of these ontologies only focus on certain domains and language rather than the contents of the Quran [3]. An ontology is an explicit specification of a conceptualization [4] i.e., to represent a common structure of information that will help human and machine to find information and return the precise answer to user queries. Ontology is the semantic search backbone, and the success of any semantic application depends on the design and development of the Ontology [5]. This paper describes the extraction of the knowledge from the Tafsir performed by Muslim Scholars that enables users to search and retrieve relevant information based on the Quran ontology that includes several fields like economy, health and others.
Section 2 presents an overview of the Quran and ontology. The proposed architecture of the Quranic search engine is described in Section 3. Section 4 provides experimental results of the keyword search and domain extraction. Finally, a conclusion in Section 5.

2. Background

This section provides an introduction to the Holy Quran and the ontologies. The Quran acts as a practical guide of life to Muslims in the personal and official spheres of life. To make it more accessible to different nations and peoples of various countries, it has been translated into almost every language of the world.

2.1. The Holy Quran

The Holy Quran contains Allah messages (words of God), which was revealed to Prophet Muhammad S.A.W. It is widely regarded as the finest piece of literature in the Arabic language and the central religious text of Islam. The Quran consists of 30 Juzs

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(parts), 114 surahs (chapters) and approximately 6236 verses (ayahs). The first surah revealed to Prophet Muhammad S.A.W is Surah Al-Alaq, the longest surah in the Quran is Surah Al-Baqarah, which consists of 286 verses, and the shortest surah is Al-Kauthar with 3 verses. Three-fourth (86 out of 114) of the Surahs were revealed during the 13 years of the Prophet's mission in Makkah; the remaining 28 were revealed during the entire 10 years of his life in Madinah. The contents of the Quran cover all aspects of our life and usually referred as a book of knowledge. Lately, scientists have proven that the Quran contains directives to modern science in various fields. However, the knowledge of science in the Quran is not yet known and explored by human beings. This will be a big opportunity for Muslims to lead in the discovery and exploration of science in the future [6]. Therefore, it is a golden opportunity for the Muslims to explore new science knowledge in the Quran. Usually Quran and Hadith are being referred in the area of Islamic laws and anything that related to religion alone, and no more than that. While this revelation covers all mysteries and secrets of science knowledge, which can only be, discovered by scientists who are experts in the intricacies of the Quran and Hadith or with the collaboration between scientist and expert in the field of Quran and Hadith.

2.2. Ontology

Ontology is the semantic search backbone, and the success of any semantic application depends on the design and development of the ontology [2]. In general, there are 2 types of ontology, the domain specific ontology and upper ontology. The domain specific ontology represents the particular meaning of the term as they apply to that domain. For example, the word "bank" has many different meanings. Ontology to represent the domain of nature would model "river bank" as the meaning of the word, while ontology about the domain of economy would model the "bank balance sheet" and "bank notes". Upper ontology is a representation of the common concepts that are applicable via a wide range of domain ontologies [7]. There are only several ontologies that have been developed based on the Quran. OntoSelect Ontology library shows that there is a lack of Arabic Ontologies Library, and about 49% of ontologies are created for the Latin character set [8]. Since each language has its own linguistic environment and cultured context, so each language needs its own ontology. The most common ontologies for English are Upper Model, Wordnet, Sumo and OpenCyc while most common work is Arabic Wordnet [9]. Up until 2017, several ontologies have been developed for Quran are focusing on certain themes or domains. Basir et al. 2017 [10] provides recent examples of research involving Quranic retrieval, as well as, examples on Quranic ontologies. Details on the quranic retrieval methods can also be found in [10]. Based on the ontology review in [10], most ontology built for the Quran only focus on a specific domain and lack in feeding cross-domain representation. Therefore, this project aims to produce a heterogeneous ontology that covers all knowledge in the Quran.

3. Quranic Ontology Examples

Various works on Quranic Ontology have been published recently [5, 11-16]. Language Research Group University of Leeds [10, 14] has developed 'The Quranic Arabic Corpus' which is focused on linguistic resource based on Arabic grammar, syntax and morphology for each word in the Holy Quran, however, a description of the concepts are insufficient especially in performing semantic search [15]. In [17] has introduced semantic-based retrieval system for the Arabic text, which will expand input query semantically for Arabic domain ontology. In [11] constructed an ontology that related to prayer in the Quran, and in [12] developed an ontology of semantic lexicons of Arabic that focuses on verses related to 'time'. In [5] produced an ontology that covers only verses in Juz' Amma. In [13] developed an ontology for living creatures including animals and birds mentioned in the Holy Quran. Another work has been done by [16] using lexicon-syntactic patterns where it gives poor performance on ambiguous Arabic text.

Based on the above review, most ontology built for the Quran only focus on a specific domain and lack in feeding cross-domain representation. Here, our work differs in term of the extraction of the knowledge in the Holy Quran. Our ontology named as 'Quranic Ontology' is developed based on the Tafsir performed by Muslim Scholars that enables users to search and retrieve relevant information based on the desired keyword (as a single word or multiple phrases with logical operator), search weightage (rank) or keyword expansion from the domain knowledge that includes several fields like economy, health and others (http://www.usimquranic.my/). Our aim is to produce a heterogeneous ontology that covers all knowledge in the Quran.

4. Quranic Search Engine Architecture

Figure 2 depicts the architecture of the Quranic Search Engine. There are two phases involved, which are the USIM's quranic ontology development and the query process. The objective of this architecture is to improve the extracting capability of the search engine from the USIM's Quranic Ontology.

4.1. Phase 1: USIM's Quranic Ontology Development

The objective of this phase is to build USIM's Quranic Ontology. This phase involved extracting relevant terms (as concepts), the relationship between concepts and related field likes economy, medical and health from the Quran by Muslim Scholars.

In the evaluation process, the extracted concepts, relationships and field are evaluated by domain experts. The experts are selected based on their expertise in Economic domain, Medical and Health domain and Quranic Tafsir domain. Figure 3 depicts the USIM's Quranic Ontology concepts that are presented in the Protege ontology editor and TopBraid semantic web modeling tool. In this figure, three core concepts (also called classes) have been defined: (1) Quran, (2) Economics, Business and Management, and (3) Medical and Health Science. These concepts are based on Malaysian Research Development Classification System 5th Edition such as nursing, clinical psychology, pharmacy, anatomy, radiotherapy and etc. [6]. The Quran has three significant sub-classes: (1) Ayat, (2) Juz and (3) Surah. For example, the first juz of the Quran is the surah al-Fatihah. This surah has seven verses.

Object properties define the relationships between two classes. The following object properties have been established: belongToJuz to indicate the Ayat to belong to which Juz, belongToSurah to indicate the Ayat to belong to which Surah, hasRelatedAyat to indicate the field (instance) that is related to any Ayat. For example, immunology hasRelatedAyat al-Baqarah_13, which means the field immunology is related to surah al-Baqarah ayat 13.

Data properties link concepts to data values (like a surah number to a surah). The following data properties have been established: hasAyatImlaie to store the Imlaie Arabic Quran Verse, hasAyatUthmani to store the Uthmani Arabic Quran Verse, hasAyatNumber to indicate the versus's number, hasJustification to indicate the verses justification, hasKeyword to indicate the keyword of the individual, hasSurahNumber to indicate the number of the Surah, hasTotalAyat to indicate the total number of verses in each surah, hasTranslation to indicate the translation of the verses, and hasDescription to indicate the description of the instance (if any). For example, searching for "food" as the keyword will provide the Surah al-Baqarah ayat 58 as one of the results. Thus, provide the following properties: food hasAyatImlaie

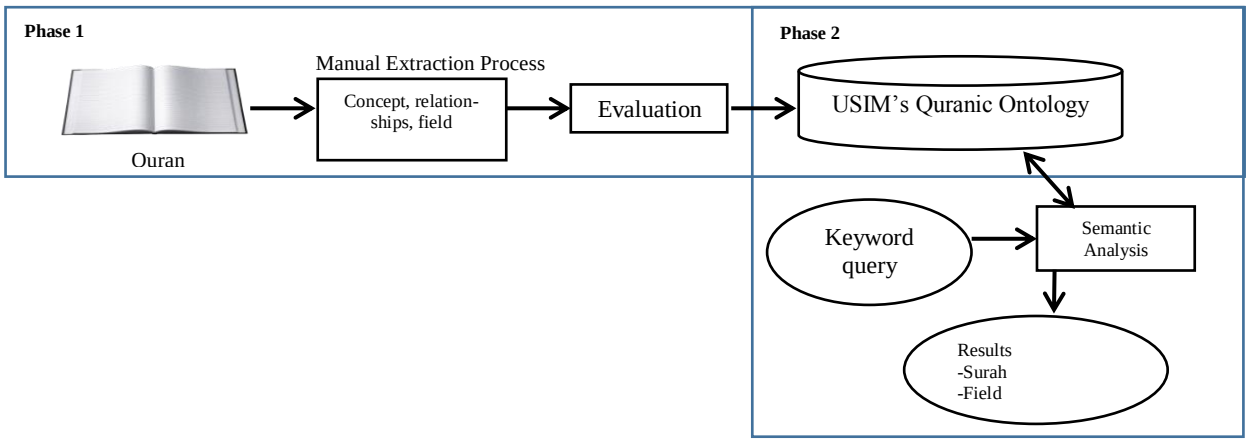


Fig. 2: System Architecture

وَإِذْ قُلْنَا ادْخُلُوا هَٰذِهِ الْقَرْيَةَ فَكُلُوا مِنْهَا حَيْثُ شِئْتُمْ رَغَدًا وَادْخُلُوا الْبَابَ سُجَّدًا وَقُولُوا حِطَّةٌ نَغْفِرْ لَكُمْ خَطَايَاكُمْ ۚ وَسَنَزِيدُ الْمُحْسِنِينَ

Al-Baqarah_ayat_58 hasAyatNumber 58, Al-Baqarah_ayat_58 hasTranslation "And when We said: Enter this city, then eat from it a plenteous (food) wherever you wish, and enter the gate making obeisance, and say, forgiveness. We will forgive you your wrongs and give more to those who do good (to others)".

4.2. Phase 2: Query Process

The keyword entered by the user is treated as an input to the Quranic Search Engine. The system will automatically extract the relevant verses and related domain for each extracted verse from the USIM's Quranic Ontology.

Figure 4 shows the ontology information retrieval alignment. Here, the similarity of the query is searched and matched based on the language-based, structured–based and string-based techniques. Element-level matching technique relies on ontology entities (or instances) and disregarding relations with other entities (or instances of others). The string-based similarity measures and identifies correspondence based on name and language-based technique identifies correspondence using NLP and lexical resources to capture conceptual similarity. On the other hand, the structure-level technique analyzes how entities (or their instances) appear together in a structure.

For example, economic *hasRelatedAyat al-Baqarah_267*, which means the field economic is related to surah al-Baqarah *ayat* 267. Figure 5 depicts the graphical representation of triples. In this figure, *spend* and *al-Baqarah_267* are concepts and *belongToSurah* is a predicate to show the relation between spend and *al-Baqarah_267*. This system also will present verses in the Quran and the translation of verses to the user. Figure 6 demonstrates an example of a keyword search. Typing in the keyword "spend" will retrieve all Ayat, Juz and Surah in the Quran, which contains the term "spend". Figure 7 depicts an example of further information displayed when one of the search result links is clicked.

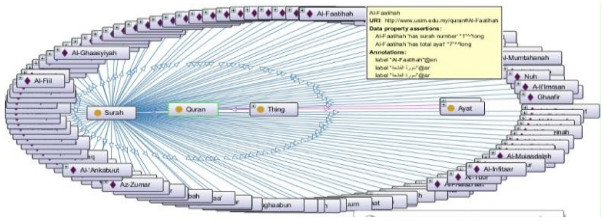


Fig. 3: Quran Ontology Concepts

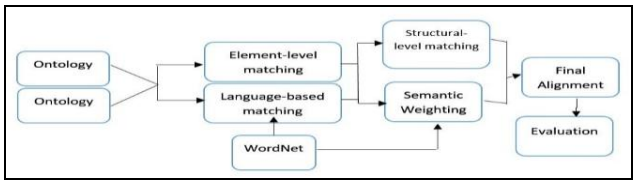


Fig. 4: Ontology Information Retrieval Alignment

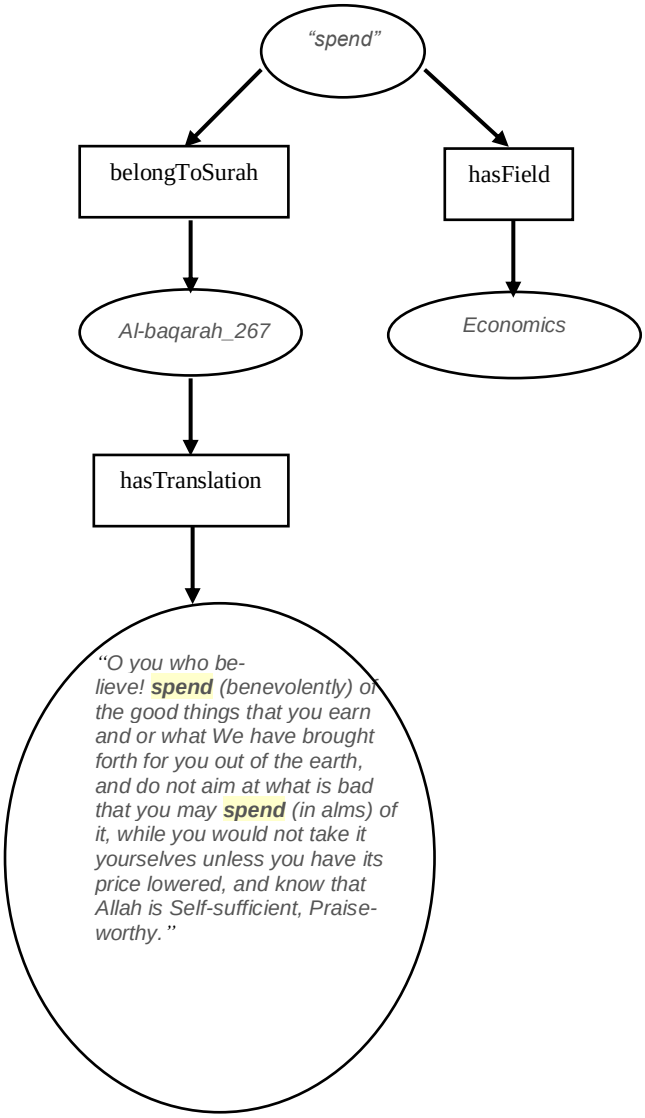


Fig. 5: The Annotation Concepts

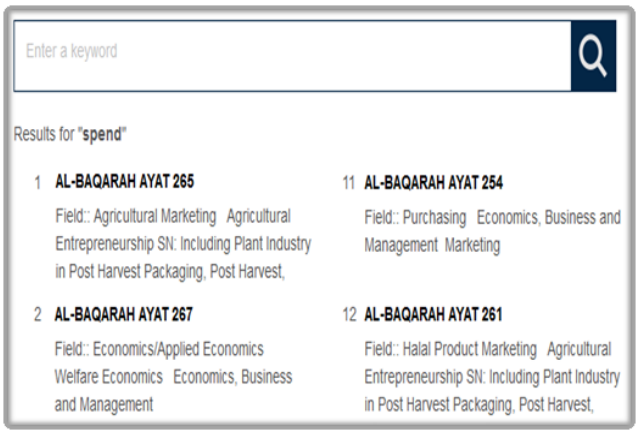


Fig. 6: Example of a keyword search

5. Experiment: Versed and Domain Extraction

The aim of this experiment is to extract related verses and domains that match the keyword entered by the user. Given a keyword query (i.e. A keyword or a list of keywords), the goal is to find the verses and domains that are related to the keyword. Here, the Quranic Ontology is developed based on the Tafsir performed by Muslim Scholars on the Quran related to several fields like economy, medical and health.

To elaborate the implementation of the USIM's Quranic Ontology, a good example of a keyword query is "disease", since it has various meanings, for instance, "disease" has conjunctive relation to the keyword "ill health". The keyword 'disease' can be found in verse 10 of Al-Baqarah, verse 52 of Al-Maaidah, and etc. On the other hand, the keyword 'ill health' can be found in verse 128 of An-Nisaa, verse 43 of Al-A'raaf, and etc. All verses found in this process will be classified as relevant.

The proposed USIM's Quranic Ontology has been experimented using six queries as depicted in Table 1. Results show that for the keyword disease, eight verses have been retrieved by the system and this keyword is related to the medical field. While, for the keyword fruit, seven verses have been retrieved by the system and this keyword is identified related to the economics, agricultural marketing, pediatrics, cardiology, vitamins, logistic, and nutrition field. For the verses retrieved for keywords such as wealth, water, spend and household can be found in Table 1.

Table 1: Relevant verses retrieved for the query

Keyword Query	Verses Retrieved
Spend	Al-Baqarah(ayat:267), Al-Baqarah(ayat:195), Al-Baqarah(ayat:219), Al-Baqarah(ayat:254), Al-Baqarah(ayat:261), Al-Baqarah(ayat:262), Al-Baqarah(ayat:265), An-Nisaa' (ayat:34), At-Taghaabun(ayat:16), Taha (ayat 6), At-Taubah (ayat:99), At-Taubah (ayat: 92), At-Taubah (ayat: 34), Saba' (ayat:39), Ibrahim (ayat:31), Al-Furqan (ayat:67)
Water	Muhammad (ayat:15), Al-Baqarah (ayat:164), An-Nisaa'(ayat:43), Al-Maaidah (ayat:6), Al-An'amm (ayat:70), Al-An'aam (ayat:99), Al-A'raaf (ayat: 57), Al-A'raaf (ayat: 160), Al-Baqarah (ayat:74)
Wealth	A-li'imraan(ayat: 186), At-Taubah (ayat:69), Al-Mujaadala(ayat:17), Al-Fajr(ayat:20)
Fruit	Al-An'aam (ayat:99), Al-An'aan (ayat:141), Al-Baqarah (ayat:265), Ibrahim (ayat: 37), Yaasin (ayat: 35), Ad-Dukhaan (ayat: 55), At-Tuur (ayat:22)
Disease	Al-Baqarah (ayat:10), Al-Maaidah (ayat: 52), Al-Anfaal (ayat:49), Al-Ahzaab (ayat: 32), Muhammad (ayat: 20), Muhammad (ayat: 29), An-Nisaa (ayat:128), Al-A'raaf (ayat:43)
Household	Al-Nahl (ayat: 80)



Fig. 7: Example of data retrieved based on searched keyword

Figure 8 depicts the Venn diagram for keywords "spend", "water", "wealth", "fruit", "disease", "household" and their corresponding domains. Based on the Venn diagram, it shows that there are several verses belong to economic or medical but some verses appear both in the economic and medical domain.

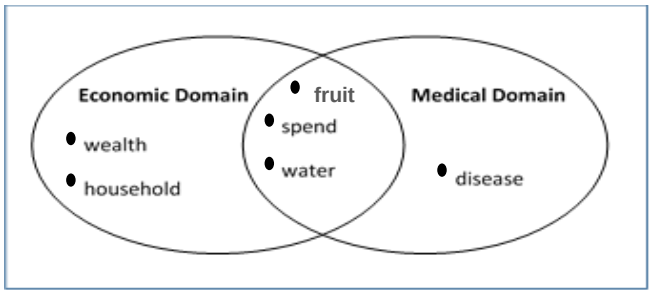


Fig. 8: The Venn diagram for the words "spend", "water", "wealth", "fruit", "disease", "household" and their corresponding domains.

For example, keyword "spend" exists in both economic and medical domains, while some keywords only exist in one domain, for example, "wealth" exists in economic domain and "disease" exists in the medical domain.

We made an analysis of our test keywords. We found that there is a possibility of one keyword to appear in various domains, and there is also a possibility for a verse to be referred to several different domains. Therefore, this search result has satisfied the following properties [18]:

- Relatedness: Each of the selected domain is related to the keyword query.
- Informativeness: The selected verses are informative enough such that they can reflect different aspects of the query.

The capabilities and benefits of the USIM's Quranic ontology are the Ummah (people) can use this reference as a guidance to enhance knowledge in a particular domain, as well as, any domain experts can use this Quranic ontology to share and annotate information in their fields according to the knowledge in the Quran.

6. Conclusion

This paper describes the extraction process of the Quranic Search Engine. The Quranic ontology construction in this search engine is

developed based on the Tafsir performed by Muslim Scholars on the Quran related to several fields namely economy, medical and health. This quranic ontology can be extended easily for other domain knowledge, as the structure of the ontology is generic. At present, the ontology focuses on the following set of data: (1) Economics, Business and Management, and (2) Medical and Health Science. The knowledge extracted from the Quran is presented as components of ontology such as the concept, relation and rules based. The results can be retrieved based on search of the desired keyword as a single word or multiple phrases with logical operators. However, there are several limitations which can be improved for future work. The existing search engine only allows keyword search (i.e., keyword → verses relation) but unable to search from verse to domain (verse → domain) or domain to verses (domain → verses). The keyword based search alone is not enough to provide accurate and relevant results. It requires a combination with semantics and synonyms that refer to the Wordnet for better results.

Acknowledgement

This work was supported by Grants: PPP/UCG-0114/FQS/30/11714, PPP/USG-0116/FST/30/11616 and USIM/NRGS_P4/ISI/8404/52113.

References

[1] N. Anas, E. A. Z. E. Alwi, M. H. Razali, R. N. Subki, N. A. A. Bakar, The integration of knowledge in Islam: Concept and challenges, Global Journal of Human Social Science Linguistics and Education, 2013, 13(10), 1-6.

[2] H. Yahya, The evolution deceit: The scientific collapse of Darwinism and its ideological background. Global Yayincilik, 2001.

[3] S. Saad, Ontology Learning And Population Techniques For English Extended Quranic Translation Text. PhD thesis, Universiti Teknologi Malaysia, 2014.

[4] N. Guarion, D. Oberle, S. Staab, What is an ontology? In S. Staab, & R. Studer (Eds.), Handbook on Ontologies-International Handbooks on Information Systems. Berlin: Springer, 2013, pp. 1–17.

[5] R. Iqbal, M. Aida, M. Y. Zulkifli, An experience of developing Quran ontology with contextual information supportnull. Multicult. Educ. Technol. J., 2013, 7(4), 333–343.

[6] K. Dukes, The Quranic Arabic corpus. Language Research Group University of Leeds 2009. http://corpus.quran.com/documentation/.

[7] M. Jarrar, Building a formal Arabic ontology (invited paper). Proceedings of the Expert. Meet. Arab. Ontol. Semant. Networks, 2011.

[8] S. Zaidi, M-T. Laskri, Review of Arabic textual terminology tools for ontologies building. Proceedings of the Management International Conference, 2009, pp. 25-28.

[9] I. F. Moawad, M. Abdeen, M. M. Aref, Ontology-based architecture for an Arabic semantic search engine. Proceedings of the Conference 10th Conference on Language Engineering, 2010, pp. 15-16.

[10] N. Basir, N.F. Nabila, N. J. M. Zaizi, M. M. Saudi, F. H. M. Ridzuan, S. A. Pitchay, Intelligent Quranic ontology retrieval. Advanced Science Letters, 2017, 23(5), 4449-4453.

[11] S. Saad, N. Salim, H. Zainal, H. Muda, A process for building domain ontology: An experience in developing Solat ontology. Proceedings of the IEEE International Conference on Electrical Engineering and Informatics, 2011, pp. 1-5.

[12] M. Al-Yahya, H. Al-Khalifa, An ontological model for representing semantic lexicons: An application on time nouns in the Holy Quran. Arab. J. Sci. Eng., 2010, 35(2), 21–35.

[13] H. Ullah Khan, S. Muhammad Saqlain, M. Shoaib, M. Sher, Ontology based semantic search in Holy Quran. Int. J. Futur. Comput. Commun., 2013, 2(6), 570–575.

[14] N. Abbas, Quran 'search for a concept' tool and website. Master thesis, University of Leeds, 2009.

[15] A. R. Yauri, R. A. Kadir, A. Azman, M. A. A. Murad, Quranic verse extraction base on concepts using OWL-DL ontology. Res. J. Appl. Sci. Eng. Technol., 2013, 6(23), 4492–4498.

[16] M. G. H. Al Zamil, Q. Al-Radaideh, Automatic extraction of ontological relations from Arabic text. J. King Saud Univ. - Comput. Inf. Sci., 2014, 26(4), 462–472.

[17] W. Alromima, I. F. Moawad, R. Elgohary, M. Aref, Ontology-based query expansion for Arabic text retrieval. International Journal of Advanced Computer Science and Applications, 2016, 7(8), .

[18] C. Muller, I. Gurevych, A study on the semantic relat-edness of query and document terms in information retrieval. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2009, pp. 1338–1347.