Integrated Modalities Search Framework for Digital English Language Text Holy Qur'an and Tafsir

Umar Bin Qushem[a], Ahmad Hafiz[b], Akram Zeki[c], Adamu Abu bakar[d]

Faculty of Information and Communication Technology
International Islamic University Malaysia (IIUM)
Kuala Lumpur, Malaysia
[a]umar.overseas14@gmail.com, [c]akramzeki@iium.edu.my, [d]adamu@iium.edu.my

Abstract

English text is commonly used online for English language and other languages as well. This is one of its attributes that makes its usage gains more popularity than others. English text version of Holy Qur'an and Tafsir can be readable by those who cannot read the Arabic version in the transliterated form. However, a non-Arab native recitation might be different from the Arab native. Nevertheless, both Arab and non-Arab require the knowledge of Tafsir for understanding Qurán word-by-word and its applications. Due to the availability of transliterated English text version of Qurán, it is possible to further utilize the Text processing aspect of generic online search engine architecture to further deliver the knowledge of Qurán. This research uses qualitative explorative approaches to present a framework through which "meaning", "teaching", "civilization", "scientific miracles" and "Aqidah and Ibadah" are integrated for any English transliterated term of Qurán into English explanation source through the knowledge of Tafsir. As a result, online user's query will extract desirable sequences of refined integration of Quran explanation inculcating knowledge composition of Holy Qur'an and Tafsir. Three short surahs of Holy Qur'an and Tafsir are used for the evaluation of the proposed framework. Text processing procedure is applied and a corpus of the English transliterated terms explanation by Tafsir of Holy Qur'an and Tafsir are generated. This framework is suitable for the design of search engine suitable for spreading the knowledge of Holy Qur'an and Tafsir.

Keywords: Transliteration, Teaching, Civilization, Scientific miracles, Aqidah, Ibadah

1. INTRODUCTION

There are many models for the application of language translation within linguistic theory, which are tied to acquisition of linguistic knowledge of different classes. Understanding the way in which a word is related to other words in the lexicon varies from language to language (Pustejovsky, 1995). Evans (2005) argued that meaning is not a property of words, but rather of the utterance: that is, a function of situated use. Those words, as such, don't have 'meanings'. The representational aspects of language that contribute to meaning involve two dimensions: lexical representations, including access to non-linguistic, conceptual knowledge, and a cognitively-realistic account of compositionality. This is the case with English language, Arabic language also is tied with important rules. However, Holy Qur'an and Tafsir, which was revealed in Arabic requires the knowledge of Tafsir for understanding. Even Arab natives require the knowledge of Tafsir for understanding the words of Qur'an and remains as an awakening issue (McLarney, 2015). English transliterated version of Holy Qur'an and Tafsir is readable as same as the Arabic version but a non-Arab native recitation

might be different from the Arab native, although both Arab and non-Arab need the knowledge of Tafsir for understanding Qurán word-by-word.

There are many online sources of Holy Qur'an and Tafsir and a lot of other Islamic articles and reports are available in different websites. Even, comprehensive online portal and mobile friendly Qur'an applications are available (Adhoni et al., 2013). Search engines are capable of returning any related English transliterated term of Quran, different from text mining, which involves extracting desirable text from text documents (Feldman & Dagan, 1995). Those transliterated terms could be further be elaborated for better understanding of the word of Holy Qur'an and Tafsir (Adhoni et al., 2014). Unfortunately, the application and explanation of the English transliterated term of Quran found online do not express in particular civilization and scientific miracle attached to the knowledge of Tafsir. Noordin and Othman (2006) found that inculcating civilization from the knowledge of Quran is suitable for the understating of its application. Although civilization is understood in different ways, some like Freud and Strachey (2005) understated it as when human tends to understand how to "use tools, and the gaining of power over fire, and the construction of dwellings", which now metamorphoses to a state of exploiting the earth for man. These are the issues, which need to be clarified by the word of Holy Qur'an and Tafsir for the public domain. Therefore, taking advantage of online search engine designed. This research uses qualitative explorative approaches to undertake a study for developing a framework that inculcates "meaning", "teaching", "civilization", "scientific miracles" and "Aqidah and Ibadah" from the words of Holy Qur'an and Tafsir to deliver into English explanation source through the knowledge of Tafsir.

2. METHODOLOGY

2. 1 Conceptual Framework

Conceptualization involves the act of processing a concept by refining its application to a conceptual or theoretical definition (Nguyen et al., 2014). The present research utilizes modalities as the core concept and aims at proposing search approaches from different points of view in order to come up with an integrated framework. The proposed conceptualization of this research is presented in Figure 1. This framework comes from evaluating some previous work related to using Holy Qur'an and Tafsir online. Elhadj (2010) proposed an online learning environment of the Quran and its sciences, which can integrate different sub-systems devoted to teaching how to recite the noble Quran and to memorize it in a manner similar to the usual way followed in the Quranic schools and circles at mosques. This is quite an interesting proposal that will enhance an engagement to Islam online and improve individual's Islamic knowledge. But, there could be some issues related to the technique used for processing and retrieval of Quranic verse in the form of mismatch or errors of reporting as observed by Sutan et al (2011). In a similar approach, Bakar and Rahman (2003) found out that Malay Translated Al-Quran documents combined search of stemming and thesaurus improve retrieval effectiveness and recall but decrease its precision. An approach to deal with this kind of problem was proposed by Sharaf and Atwell (2012). The researcher built large corpus of original Quranic Roman text transliterated version where semantically similar or related verses are linked together in the corpus. This draws the need to start building up Qur'anic search engine architecture. An attempt to develop a full functional search engine architecture for Qur'an was presented by Beirade (2014). The author's techniques dwell on the generic automatic processing of language where a database of the Qur'anic terms with Othmani writing as the backbone. Three retrieval approaches in sequences were set and

started with morphological analysis, followed by syntactic, and finally semantic analysis. As a result, the research formulated its search strategies to use the "meaning", not the word structure. This means that for any keyword query of the Qur'an, the result should be the meaning of that term retrieved from words' trees, subject's trees, and syntactic analysis trees. There are some drawbacks observed by some researchers regarding the legacy keyword searching techniques applied to Qur'an in retrieving semantically relevant verses. An attempt to address this drawback was made by Shoaib et al (2009). The researchers claimed conceptualizing retrieval strategies by extracting relevant verses from any search keyword followed by all its synonyms in an order of most relevant verses first and the least at the end. Another concept of ontology as applied to semantic web for semantic search in Quran was proposed by Khan et al (2013). The researchers suggested the adding living creatures like animals and birds mentioned in Quran for any retrieval terms related to keyword search. Yauri, et al (2013) proposed the implementation of Quran ontology to consist of noun concepts, which the researchers were able to identify in Qur'an. The researcher outlined the relationship that exists between the concepts. The retrieval techniques were designed to reformulate or pre-process query to match the content Quranic verse related to the query keyword. A diacritic-less searching approach to retrieve from the Quran relevant verses that match a user's query through automatic query expansion techniques has been proposed by Hammon et al., (2007).

Consistent with the previous work, this research proposed retrieval architecture for searching, but it also observed that both semantic and text-based retrieval approaches proposed for Holy Qur'an and Tafsir still need an enhancement. Based on that, this research proposes the use of Quranic verses alongside with Hadith for any query keyword search. Thereafter the result should include the followings:

  i.   Meaning (M) 1
  ii. Teaching (T) 2
  iii. Civilization (C) 3
  iv. Aqidah and Ibadah (A) 4
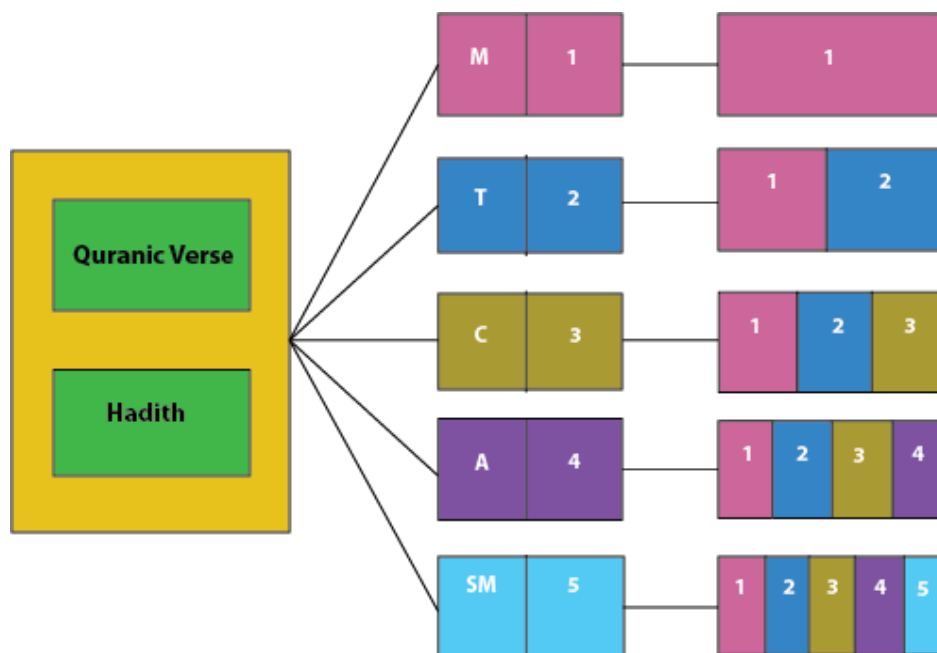 v.   Scientific miracles (SM) 5

Fig 1. Conceptual Framework

2.2 Case Study

The proposed framework was used in a case study to demonstrate the techniques for effective utilization of knowledge in public atmosphere. Three short Surahs of the Quran ('Surah An-Nusr', 'Surah Al-Quraysh' and 'Surah Al-Fill') were used in this case study. The technique involves building up a corpus where any verse of Quran will be taken into account from the user, specifically the three short surahs stated above. The corpus is logically presented by taking into account a matrix of 1, 2, 3, 4, and 5. The combination of either two or more can be retrieved depending on the query keyword used by the users. If the word contains 1, 2, 3, 4, and 5, then the result of the query can come in attached format together. Otherwise it could either be (1), or (1, 2) or (1, 2, 3) or (1, 2, 3, 4) or the combination of any of the required output. The matrix is demonstrated in Table 1.

Table 1: Matrix of query terms' results in a corpus

NO		1	2	3	4	5
		M	T	C	A	SM
1	M	1	0	0	0	0
2	T	1	1	0	0	0
3	C	1	1	1	0	0
4	A	1	1	1	1	0
5	SM	1	1	1	1	1

Using this approach, Query process can integrate 'hadith' related to that verse of the Quran where the query terms occur. Afterwards, many classes of Query suggested for the user which depend on the user needs are also attached. Subsequently, some aspects of the Quran, which make people understand clearly about the divine books, such as Meaning, Teachings, Civilization, 'Aqidah' and 'Ibadah', and scientific miracle can be acquired.

2.3 Evaluation

The evaluation for the proposed approach considers a class query where any search query term must have a meaning retrieved. Search query term could also be expressed with teaching after its meaning has been presented through the act of Tafsir with reference. In a more elaborated fashion, search term can contain classes for its corresponding civilization. The result of search query term can also be applied to its Aqidah and Ibadah and scientific miracles where available. The sequences of the output are represented below:

Class 1= Meaning
Class 2= Meaning + Teaching
Class 3= Meaning + Teaching + Civilization
Class 4= Meaning + Teaching + Civilization + Aqidah and Ibadah
Class 5= Meaning + Teaching + Civilization + Aqidah and Ibadah + Scientific Miracle

Following these classes of queries, integrated modes of retrieval can be developed. This research gathered information and searched through the Quran and also 'Tafsir' of Surah An-Nasr, Surat Quraysh, and Surat Al-Fill. The Roman English transliterated Holy Qur'an and

Tafsir version is used. The corpus is built according to term's meaning, teaching, history and civilization, 'aqidah' and significance, as well as their scientific miracles. Based on this, surah is considered as a document, thus the following documents were used:

> **Document 1** *(Surat An-Naṣr):*
> Itha Ja'a Nasr Allahi wa alfath* wa rai't alnnas yadkhloon fe deen Allhi afwaja* fa sabbeh bhamde rabbic wa istaghfirhu innhu kan tawwaba.
>
> **Document 2** *(Surat Quraysh)*:
> Le'ilaf quraysh* ilafehim rehlata Alshita'a wa alsaif* falya'abdo rabb hatha albeit*althi ata'mahum men joua wa a'amnahum men khowf.
>
> **Document 3** *(Surat Al-Fīl):*
> Alam tra kaif fa'al rabbuka be ashab alfil* alam yaja'al haidahum fi tathleel* wa arsal alaihem tairan ababeel* tarmeehum behijaraten men sejjeel* faja'alahum ka'asefen ma'akul.
>
> **Document 4** (Verse#11 Surah#44):
> Thumma istwa ila alsama'a wahia dukhanun faqal laha wa lilardh ietya tawan qalata Atyna taean.

Table 2: Surat An-Naṣr

	Verse	Meaning	Teaching	Civilization	Aqida'ah & Ibadah	Scientific Miracles
1234	إذا جاء نصرالله والفتح Itha Ja'a Nasr Allahi wa alfath	When the victory of Allah has come and the conquest,	…ask Him forgiveness …	…eternal lesson…	Duty	
1	ورأيت الناس يدخلون في دين الله أفواجا wa rai't alnnas yadkhloon fe deen Allhi afwaja	And you see the people entering into the religion of Allah in multitudes				
1	فسبح بحمد ربك و استغفره إنه كان توابا fa sabbeh bhamde rabbic wa istaghfirhu innhu kan tawwaba.	Then exalt [Him] with praise of your Lord and ask forgiveness of Him. Indeed, He is ever Accepting of repentance.				

Sbabul Nuzul:
When Allah's succour and the triumph cometh...) [110:1-3]. This was revealed when the Prophet, Allah bless him and give him peace, left from the Battle of Hunayn. The Prophet, Allah bless him and give him peace, lived only two years after this battle. Sa'id ibn Muhammad al-Mu'adhdhin informed us> Abu 'Umar ibn Abi Ja'far al-Muqri'> al-Hasan ibn Sufyan> 'Abd al-'Azizi ibn Sallam> Ishaq ibn 'Abd Allah ibn Kaysan> his father> 'Ikrimah> Ibn 'Abbas who said: "When the Messenger of Allah, Allah bless him and give him peace, returned from the Battle of Hunayn and Allah, exalted is He, revealed (When Allah's succour and the triumph cometh), he said: 'O 'Ali ibn Abi Talib! O Fatimah! Allah's succour and triumph has come, I have seen people entering the religion of Allah in troops, I therefore hymn the praises of my Lord and seek forgiveness of Him, for He is ever ready to show mercy' ".

Table 3: Surat Quraysh

code	Verse	Meaning	Teaching	Civilization	Aqida'ah & Ibadah	Scientific Miracles
1234	ألم تر كيف فعل ربك بأصحاب الفيل Alam tra kaif fa'al rabbuka be ashab alfil	Have you not considered, [O Muhammad], how your Lord dealt with the companions of the elephant?	… destroy the Ka'bah,..	…Invasion…	only Allah's inflicting His punishment,	
123	ألم يجعل كيدهم في تظليل alam yaja'al haidahum fi tathleel	Did He not make their plan into misguidance?	…plot go astray? …			
123	و أرسل عليهم طيرا أبابيل wa arsal alaihem tairan ababeel	And He sent against them birds in flocks,	… And sent against them birds, in flocks, …	… …	... …	
123	ترميهم بحجارة من سجيل tarmeehum behijaraten men sejjeel	Striking them with stones of hard clay,	… Striking them with stones of Sijjil. …			

123	فجعلهم كعصف مأكول faja'alahum ka'asefen ma'akul	And He made them like eaten straw.	… And made them like an empty field of stalks (of which the corn has been eaten up by cattle)…			
	sbabul Nuzul: *(Hast thou not seen how thy Lord dealt with the owners of the Elephant?...) [105:1-5]. This was revealed about the story of the owners of the elephant who intended to destroy the Ka'bah and how Allah, exalted is He, dealt with them by destroying them and driving them away from the Sanctuary. This story is well known.*					

Table 4 Surat Al-Fīl

code	Verse	Meaning	Teaching	Civilization	Aqida'ah & Ibadah	Scientific Miracles
1234	ألم تر كيف فعل ربك بأصحاب الفيل Alam tra kaif fa'al rabbuka be ashab alfil	Have you not considered, [O Muhammad], how your Lord dealt with the companions of the elephant?	… destroy the Ka'bah,..	…Invasion…	only Allah's inflicting His punishment,	
123	ألم يجعل كيدهم في تظليل alam yaja'al haidahum fi tathleel	Did He not make their plan into misguidance?	…plot go astray? …			
123	و أرسل عليهم طيرا أبابيل wa arsal alaihem tairan ababeel	And He sent against them birds in flocks,	… And sent against them birds, in flocks, …	… …	... …	
123	ترميهم بحجارة من سجيل tarmeehum behijaraten men sejjeel	Striking them with stones of hard clay,	… Striking them with stones of Sijjil. …			

123	فجعلهم كعصف مأكول faja'alahum ka'asefen ma'akul	And He made them like eaten straw.	… And made them like an empty field of stalks (of which the corn has been eaten up by cattle)…			
	sbabul Nuzul: *(Hast thou not seen how thy Lord dealt with the owners of the Elephant?...) [105:1-5]. This was revealed about the story of the owners of the elephant who intended to destroy the Ka'bah and how Allah, exalted is He, dealt with them by destroying them and driving them away from the Sanctuary. This story is well known.*					

Table 5: Verse#11 Surah#44

code	Verse	Meaning	Teaching	Civilization	Aqida'ah & Ibadah	Scientific Miracles
1234	ثم استوى إلى السماء و هي دخان فقال لها وللأرض ائتيا طوعا أو كرها قالتا أتينا طائعين. Thumma istwa ila alsama'a wahia dukhanun faqal laha wa lilardh ietya tawan qalata Atyna taean.	Then He directed Himself to the heaven while it was smoke and said to it and to the earth, "Come [into being], willingly or by compulsion." They said, "We have come willingly.	We have to belive in Allah's well.	Source of sintific knowledge.	Allah S.WT has an Istwa'a on his A'rsh and this Istwa'a as Imam Malik said:" the istwa'a is known, the how is unknown and asking about that is Bida'ah.	These verses has been recited in miracles Quranic scientific conference which was held in Cairo and when the Japanese professor (Yoshida Kozai) heared that verse got surprised and said science and scientists did not reach this amazing fact only recently after it picked up a powerful

3. Discussion

The proposed use of class query from 1 to 5 has been demonstrated in Section 4. Table 2 to 6 express the coding scheme applied. At first, the column for "meaning" is the direct meaning of a verse from Holy Qur'an and Tafsir. Any term within that verse resides in that meaning sourced from the act of Tafsir. Therefore, if a search is either for a term or verse, then the entire meaning will be the result. At this level, the class of the query is 1. That is basic unit from any search except where the search term is an error term. Following the first class query result, is the next column, which is the teaching class query term result column represented in 2. This is for verses in most cases. The reason is that hardly a term could describe a phenomenon, but in some cases where it did, it is also applied to it. At this level, the result of the query whose meaning was already tagged as 1 will be highlighted if teaches something, so that the result is code 12. This highlight should capture any keyword or phrase within the teaching that verse or term provided (see Table 2). However, very important term is recommended in order to deliver the knowledge of the verse or term faster for the user. Furthermore, this keyword that represents the teaching of the verse in some cases could be the query term. That query can come from the term. As a result, it will give the verse of the Holy Qur'an and Tafsir. In this case, a reserve engineering from "teaching to verse", which will be coded 21 will appear. This will lead to the acquisition of knowledge in a better way. The next layer is civilization and coded by 3. Similar to the teaching column, this column provides the explanation of what the verse or term applies to the civilization where applicable sources from the act of Tafsir. This means verse meaning and teaching will be integrated with civilization. The query result of verse or term with meaning, teaching, and civilization is coded by 123 and the output is also a highlight of a very important term within the details of how the verse is applied to the civilization. This will also be reserved engineered and the result could either be 321 or 312. The remaining two columns are coded 4 and 5, which bear Aqida'ah & Ibadah and Scientific Miracles, respectively and the retrieval method is also applied similar to them.

This presentation requires a lot of filtering and good application of coding scheme. The codes help in sending highlights, which subsequently lead to the exploration of detailed result. In this case, the details of class 2, 3, 4, and 5 can only be delivered when require to go further to explore. However, even at the highlight levels, this method helps in elaborating transliterated version of Holy Qur'an and Tafsir verses meaning and knowledge. Search engine exhibiting this proposed framework improves the search result quality and the teaching of Holy Qur'an and Tafsir. This technique can lead to the exposure of Holy Qur'an and Tafsir online and a lot of other Islamic articles and Islamic reports available in different websites. The combination of the codes 12345 means search engine is able to crawl and makes gathered verses or terms with all the attributes in document store. Search engine is now capable of returning any interrelated query term of Quran.

4. Conclusion

This paper proposes an integrated modalities search framework for text Holy Qur'an and Tafsir. In building a search engine architecture, text processing is crucial. When it applies to English Roman transliterated version of Holy Qur'an and Tafsir, the treatment of text processing could not be the same with the generic text processing approach. Since there are lots of independent cases in the use of English text, this does not apply to Holy Qur'an and Tafsir, because it is restricted (Sardar, 2011; Hossein, 2013). As a result, this research provides modalities for a search engine architecture, especially for quality retrieval of verse

or term related teaching, civilization where applicable. This procedure is presented by filtering and application of coding scheme. The codes help in providing highlights, which subsequently give the detailed elaboration of transliterated version of Holy Qur'an and Tafsir verses meaning and knowledge. This is possible where only English transliterated terms explanation by Tafsir are provided, otherwise no result is expected.

5. REFERENCES

Adhoni, Z. A., AhadSiddiqi, A., & El Mortaji, L. (2014). Transliteration of knowledge retrieval in Urdu from Holy Quran using state of the art information technologies. In Optimization, Reliabilty, and Information Technology (ICROIT), 2014 International Conference on (pp. 359-363). IEEE.

Adhoni, Z. A., Al Hamad, H., Siddiqi, A. A., & El Mortaji, L. (2013). Towards a Comprehensive Online Portal and Mobile Friendly Qur'an Application. In Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences (NOORIC).

Bakar, Z. A., & Rahman, N. A. (2003). Evaluating the effectiveness of thesaurus and stemming methods in retrieving Malay translated Al-Quran documents (pp. 653-662). Springer Berlin Heidelberg.

Beirade, F. (2014). Search engine for Holy Quran. In ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb), 2014 4th International Symposium (pp. 1-6). IEEE.

Elhadj, Y. O. M. (2010). E-Halagat: An E-Learning System for Teaching the Holy Quran. Turkish Online Journal of Educational Technology-TOJET, 9(1), 54-61

Feldman, R. & Dagan, I. (1995) Knowledge discovery in textual databases (KDT). In proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, August 20-21, AAAI Press, 112-117.

Freud, S., & Strachey, J. (2005). Civilization and its discontents. WW Norton & Company.

Hammo, B., Sleit, A., & El-Haj, M. (2007). Effectiveness of query expansion in searching the Holy Quran.The Second International Conference on Arabic Language Processing CITALA'07. , Rabat, Morocco, pp. 1-10.

Hossein, N.S.(2013) Islamic Spirituality: Foundations. Routledge.

Khan, H. U., Saqlain, S. M., Shoaib, M., & Sher, M. (2013). Ontology based semantic search in Holy Quran. International Journal of Future Computer and Communication, 2(6), 570-575.

McLarney, E. A. (2015). Soft Force: Women in Egypt's Islamic Awakening: Women in Egypt's Islamic Awakening. Princeton University Press.

Nguyen, Q. D., Fernandez, N., Karsenti, T., & Charlin, B. (2014). What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. Medical education, 48(12), 1176-1189.

Noordin, M. F., & Othman, R. (2006). An information retrieval system for Quranic texts: a proposed system design. In Information and Communication Technologies, 2006. ICTTA'06. 2nd (Vol. 1, pp. 1704-1709). IEEE

Pustejovsky, J., & Bouillon, P. (1995). Aspectual coercion and logical polysemy. Journal of semantics, 12(2), 133-162

Sardar, Z. (2011). Reading the Qur'an: The Contemporary Relevance of the Sacred Text of Islam. Oxford University Press

Sharaf, A. B. M., & Atwell, E. (2012). QurSim: A corpus for evaluation of relatedness in short texts. In LREC (pp. 2295-2302)

Shoaib, M., Yasin, M. N., Hikmat, U. K., Saeed, M. I., & Khiyal, M. S. H. (2009). Relational WordNet model for semantic search in Holy Quran. In Emerging Technologies, 2009. ICET 2009. International Conference on (pp. 29-34). IEEE.

Sultan, A. M., Azman, A., Kadir, R. A., & Abdullah, M. T. (2011, June). Evaluation of Quranic text retrieval system based on manually indexed topics. In Semantic Technology and Information Retrieval (STAIR), 2011 International Conference on (pp. 156-161). IEEE.

Yauri, A. R., Kadir, R. A., Azman, A., & Murad, M. A. A. (2013, March). Ontology semantic approach to extraction of knowledge from Holy Quran. In Computer Science and Information Technology (CSIT), 2013 5th International Conference on (pp. 19-23). IEEE.

Zeki, A., Oyefolahan, I. O., Malek, A., Azrin, N., Mahmud, M., Mahmood, N. & Abubakar, A. (2014). Investigating how the sources of online audio Qur'an empower listeners. In the 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M), 2014 (pp. 1-6). IEEE.