International Journal on Islamic Applications in Computer Science And Technology, Vol. 5, Issue 2, June 2017, 13-25

Retrieval Performance for Malay Quran

Nor Diana Ahmad[a], Brandon Bennett[b], Eric Atwell[c]

[a,b,c]School of Computing
University of Leeds
Leeds, United Kingdom
[a]Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450, Shah Alam, Selangor, Malaysia
[a]scnda@leeds.ac.uk, [a]nordiana@tmsk.uitm.edu.my, [b]b.bennett@leeds.ac.uk, [c]e.s.atwell@leeds.ac.uk

Abstract

Information Retrieval (IR) plays an important role in retrieving information related to the user's query. IR relies on finding relevant data from a set of knowledge resources, such as the Quran. Finding information from the Quran can be based on metadata, indexing, or other content-based methods. The Quran is the most widely read book in the world and automating knowledge retrieval from this religious book in general is very challenging. One of the challenging tasks is when the Quran has been translated into various languages around the world by Muslim experts, such as Malay. This has led to the development of a number of search applications in multiple languages, which can retrieve knowledge based on keywords. Retrieving the knowledge from the Quran using keywords has several fundamental problems. In many cases, the searching cannot retrieve the relevant concept or knowledge and verses. Therefore this research investigated the existing problems related to the study of Quran and studies the current search techniques used in research on the Quran, especially the Malay Quran. This research will help researchers understand and learn the current scenario of research on the Quran.

Keyword: The Quran; Malay Quran; Semantic Search; Ontology; Information Retrieval.

The Quran is fundamental to all Muslims because it contains comprehensive guidance to Muslims in all aspects of life. Muslims are required to read and learn not only the Quran but the meaning of Quran in languages they understand. This will help Muslims to perform their daily routine efficiently. Nowadays, people are looking for the easiest ways to learn the Quran. Many search applications have been built to facilitate the retrieval of knowledge from the Quran. Current approaches that can be exploited for searching and retrieving information from Quran use semantic-based and keyword-based techniques. The semantic-based technique is a concept-based search tool that retrieves results based on word meaning, or concept match, whereas the keyword-based technique returns results based on letters matching word(s) as queries (Alqahtani & Atwell, 2016). At the moment, most Quran search tools employ the keyword search technique.

Information Retrieval (IR) plays an important role in retrieving information related to the user's query. The creation of computer and internet has made easier and faster the task of retrieving information. Despite of that, there are still issues pertaining retrieving information. For example, if a

user types "Malay Quran" in a search engine, it will retrieve many results. Although this contains relevant information required by the user, it is hard for the user to look at all links. If a user limits the search to a specific version of Malay Quran, the user may loose a few other links that contains valuable information about the Quran (Wahid, 2014).

The study of Ullah Khan et al., (2013) highlighted that searches and retrieval of knowledge in the Quran is sometimes unclear and less accurate due to the non-implementation of dynamic ways for retrieving knowledge or verses. Furthermore, the present way of searching in the Quran is not based on knowledge classification (Ullah Khan et al., 2013). Therefore, this shows that there are weaknesses in searching and retrieving the verses in the Quran due to the non-use of appropriate techniques. In conjunction with that, this research investigates the existing problems related to the study of the Quran and studies the current search techniques used in research on the Quran, especially the Malay Quran.

2.1Information Retrieval

Information Retrieval (IR) can be defined as discovering information that meets an information need using expansive accumulations of text usually stored on computers (Manning et al., 2008). At the beginning of the emergence of IR, only a few people were involved directly with retrieval and searching process such as librarians, paralegals etc. (Baeza-Yates & Ribeiro-Neto, 2011). Presently the situation has changed and a huge number of people now consistently engage with IR when they use World Wide Web (WWW) search tools, for instance, Google or Yahoo. According to Baeza-Yates & Ribeiro-Neto (2011), the main goal of IR is to retrieve all documents relevant to a user's query, while at the same time retrieving as few non-relevant documents as possible. It is a balancing act between finding a large amount of potentially-relevant information and losing some potentially valuable information (Wahid, 2014).

2.2 The Quran

The Quran is the most widely read book in the world. The Quran has stimulated the interest of many researchers, especially in the field of information systems to assess and automate the extraction of knowledge. This led to the development of a number of search applications, which aim to provide retrieval of knowledge to facilitate people to understand people in better way. However, understanding and retrieving the knowledge from the Quran is a major research challenge for computer science and artificial intelligence (Atwell et al., 2010). The search method in Quran faces several fundamental problems, such as the inability to retrieve relevant knowledge and verses (Yauri et al., 2012). The further study proved by Tian Tian (2012) in her thesis that most search engines suffer from three common problems in Natural Language Processing such as the synonym problem, the homonym problem, and the wrong granularity problem. The synonym problem appears in the form that the user might send a different term to the search engine than what is contained in a document. For instance, a user query "the last prophet in Islam" might miss documents with Prophet Muhammad SAW, even though these two terms are synonymous.

Nowadays, the Quran has been translated into various languages around the world by Muslim experts. The main aim of the availability of the Quran translations is to allow the reader to understand the Quran in clearer ways. As for Malay readers, there are many available Malay Quran translations. Nevertheless, there are a few issues regarding the Malay Quran translations such as ambiguities of words, lack of word equivalence between Malay and Arabic or Malay with English, and different structures of word, sentence, and discourse in these two languages (Tabrizi & Mahmud, 2013). Besides, Othman & Wahid (2011) claimed that standardization of different versions of Malay Quran was challenging due to the need of alignment by meanings. Different versions of Malay Quran surfaced due to the manner translations were made, which could likely be from a secondary source for some of the texts. Standardization issues could be reduced if the translations were made from the original texts with the references to other authentic sources like Hadith or *Tafsir*.

2.2.1 Related works on Malay Quran Research

Most of the research on Malay Quran focused on Natural Language Processing area. Each research suggested different techniques with the same purpose for improving the information retrieval. Hasmy et al., (2015) proposed stemming and thesaurus to search and retrieve relevant Malay translated Qur'an documents based on user natural query words. A stemming algorithm is an automated procedure that reduces words with the same stem to a common form, usually by removing derivational and inflectional suffixes from each word. The authors stated by using stemming, the efficiency of document retrieval is increased since the size of index files is reduced by 50% as a result of grouping many morphological word variants into a single stem word. Based on the experiment, the combination of stemming and thesaurus methods retrieved 60.22% and it proved to be more effective in retrieving more relevant documents compared to other methods such as extract match, stemming, or the thesaurus.

On the other hand, Yunus, Zainuddin, & Abdullah (2010) presented Stemming Semantic Query (SSQ) as a new approach to improve the retrieval of verses for Qur'an documents results. The authors compared the semantic results and stemming semantics result using three different languages i.e. English, Malay and Arabic. This research found that the semantic approach with stemmer contributed to better performance of retrieving more relevant and related Qur'an documents. Furthermore, Yahya, Abdullah, Azman, & Kadir (2013) proposed a semantic search for the Qur'an based on Cross Language Information Retrieval (CLIR). In this research, they evaluated a CLIR approach based on domain ontology that used Quran Arabic concepts by Dukes (2013) for disambiguation of the translation of a given query and enhancing dictionary-based query translation.

One Malay Tagger is developed by Omar (2011) which applies trigram Hidden Markov Model (HMM) method to identify words tags in Malay sentences. The model is tested using a corpus of 18,135 tokens tagged with a set of 21 tags similar to the set of tags used by *Dewan Bahasa dan Pustaka* (DBP). The results show that the best predictions are made with

accuracy 67.9% using only prefixes information with a fixed prefix length equals to three letters. Similar results with accuracy 66.7% are achieved using a combination of the first and the last three letters of each word. When using suffixes information only, the best accuracy achieved is 60% with suffix length of five letters. These findings show that HMMs are suitable models to be used to predict any Malay word's POS tag. In addition, Alfred, Mujat, & Obit (2013) proposed a rule-based method for identifying Malay POS tags called RPOS. It applies affixing and word relation rules to determine the right word category. Malay words can be formed with prefixes, suffixes, circum-fixes and/or infixes. In their paper, the authors consider infixes less important and not effective for the task of POS tagging. Different affixes can be categorized in different word categories. When there is more than one possible tag for the word, word relation rules are applied to identify the most suitable POS tag based on the context. If the word is not found in the POS tag dictionary, affixing rules are applied to determine possible tags for the word and then the word relation rules are applied to solve the ambiguity (if any). The POS dictionary is manually built from Malay Thesaurus by DBP and used to assign all possible tags to each word in a Malay sentence. The results of this rule-based method show that it has higher performance than the statistical POS tagger with an accuracy of 89% for Malay news articles and 86% for Malay biomedical articles. This shows that it is able to predict unknown words' POS tags at reasonable accuracy. However, this tagger fails to tag words that are borrowed from English and also words that have no affixation especially proper nouns. Richer relation rules are needed to improve the tagging results of RPOS tagger.

Lastly, Xian et al., (2016) proposed Mi-POS, a Malay language POS tagger that was developed using a probabilistic approach with information about the context. This research used manually-built corpus contains 152 articles with a total of 64,534 tokens from *Bernama* news archive. It was manually tagged by a Malay native speaker who assigned a single POS tag to each word. Then the tagged corpus was verified by two other Malay native speakers to correct mistakes and solve ambiguity if any. There is a total of 13 non-symbols POS types used to tag the training corpus by the Malay native speakers. The authors also compared the result with other Malay POS taggers such as Lazy Man's Tag and Trigram HMM. The final results showed that Mi-POS outperforms other Malay Part-of-Speech taggers in terms of accuracy with an accuracy of 95.16% obtained by tagging new words from the same training corpus type and 81.12% for words from different corpora types.

Although there are few studies that have been done, however there are limited resources and tools that are available or made accessible for computational linguistic analysis of the Malay Quran.

2.3 Semantic Search

Semantic search is an application of Semantic web showing significant potential for improving the performance of retrieval. Semantic search is a searching method in which a search query aims not to only find keywords, but also to determine the intent and contextual meaning of the words a person is using for search. This might be achieved by adding semantic tags to the documents in order to structuralize and conceptualize the object within documents (Alqahtani & Atwell, 2016). The main goal of semantic search is to deliver information in a meaningful way rather than having to sort through lists of documents bound by loosely-related keywords.

Many desktop and web applications have been developed to retrieve knowledge from the Quran. There are two techniques used to retrieve information from Quran: semantic based and keyword-based (Alqahtani & Atwell, 2015). (Shoaib et al., 2009) proposed a relational model for semantic search in the Quran using the WordNet relationships. This relational model creates the taxonomy of the related terms in Surah Al-Baqarah (Chapter 2 from the Quran). The model facilitates performing a subject search for the Quran readers and provides a framework capable of retrieving related verses from the Quran. This model of semantic search showed 80% accuracy. However, during the retrieving process, some irrelevant verses are also retrieved. Here, the authors also discussed the problem of the current keyword based searching and the issues related to semantic search in the Quran. This research has contributed to the improvement of semantic search in the Quran. In the future, the authors intend to extend this work that can eliminate irrelevant verses. Later, the researchers proposed a new research on ontology based semantic search. Tarawneh & Al-Shawakfa (2015) presented a new hybrid method called The Baheth Searcher of the Quran text using the combination of syntactic (keyword) and semantic based approach to index and search Quran text.

Nassourou (2011) presented a methodology of reconstructing the Quran's chronology based on machine learning techniques. A hybrid statistical classifier have been employed, in order to get the plausible dates of revelation in accordance with the traditional Islamic scholars and western orientalists chronologies. Later, the author produced a new technique to categorizing the Quran chapters. Nassourou (2012) proposed a new algorithm using machine learning techniques for categorizing the chapters in the Quran. In this research, the author used SVM and naïve Bayesian as functional classifiers. The categorization model of the chapters was based on the phases of the messenger ship of Prophet Muhammad. Adhoni & Al Hamad (2013) developed a cloud-based of the Quran portal using Drupal technology. In their work, the authors build the portal which can be used to search the knowledge from the Quran in more than one language. AlMaayah et al (2014) developed WordNet for Quran by creating semantic connection between words in order to get better understanding of the meaning of Quran words. This study focused on classical Arabic of the Quran rather than the modern standard Arabic.

2.4 Ontology

Gruber defined ontology as an explicit specification of a conceptualization (Gruber, 1993). A common goal developing ontology is to share common understanding of the structure concepts between people and software agents, and permit reuse of domain knowledge (Gruber, 1993). Ontology consists of four components; concepts, relations, axioms, and instances.

Ullah Khan et al., (2013) describe ontological work for searching the concepts from the Quran using the semantic search technique. In this research, the authors proposed a new framework that describes the semantic search. This framework can be applied to any Islamic document. This research uses living creatures including animals and birds mentioned in the Quran as the sample domain. Unfortunately, this research is not very clear in explaining the final result and how the framework can improve the current ontology based semantic search in Quran. Uthayan & Mala (2015) suggested a new querying mechanism for information retrieval which integrates the ontology queries with keyword search. This research proposed a new hybrid method based on matching extracted instances from the queries and information field. Besides, Ta'a et al., (2013) developed an ontology to represent and classify the knowledge of Quran using thematic approach. Thematic or theme-based approach focuses on themes within a story to give narratives a sense of direction and purpose (Hargood, Millard, & Weal, 2008). This article focuses on the themes of the Quran as describes in *Syammil Quran Miracle the Reference*: Quran, 'Akhlak' and 'Iman'. This Quran ontology help user to understand the Quran in a systematic way. Alrehaili & Atwell (2014) reviewed the past approaches or works related to computational ontologies based on nine (9) criteria: Quran Text, Coverage Area, Coverage Proportion, Underlying Format, Underlying Technology Used, Availability, Concepts Number, Relations Type, and Verification Method Used. Table 1 below summarize these criteria;

Table 1 : Summary of criteria based on review by (Alrehaili & Atwell, 2014)

No	Criteria	Previous Work
1.	Quran Text	(Ullah Khan et al., 2013), (Saad et al., 2010) and (Ismail et al., 2015) built ontologies using English translation of Quran. (Yahya et al., 2013) and (Yunus et al., 2010) used Malay translation of Qur'an as a sources for built the ontology.
2.	Coverage Area	(Ta'a et al., 2012) covered on the themes of the Quran as describes *in Syammil Quran Miracle the Reference*: Quran, 'Akhlak' and 'Iman'. While (Ullah Khan et al., 2013) used living creatures including animals and birds mentioned in the Quran as the sample domain ontology.
3.	Coverage Proportion	Entire Quran: (Muhammad, 2012) in his PhD thesis used all the chapters in Quran. (Hakkoum & Raghay, 2015) developed the whole Quran ontology by adapting from several resources. Some parts: There was no research conducted cover the some part of Quran. Specific Topic : Most of the research covered specific topic such as (Shoaib et al., 2009) just focusing on the chapter 2 of Quran.

4.	Underlying format	(Yahya et al., 2013), (Ta'a et al., 2012) and (Ullah Khan et al., 2013) used OWL to build the ontologies. (Yauri et al., 2014) used RDF/XML format.
5.	Underlying technology used	(Ullah Khan et al., 2013) used protégé and SPARQL.
6.	Availability	(Hakkoum & Raghay, 2015) developed the Quran ontology and it can be accessed through http://quranontology.com/.
7.	Concepts number	(Dukes, Atwell, & Sharaf, 2010) defined 300 concepts in Quran.
8.	Relations Type	(Shoaib et al., 2009) showed the synonyms relations. (Saad, Salim, & Zainal, 2009) shows meronyms (*partOf*).
9.	Verification Method Used	(Saad et al., 2009) and (Saad et al., 2010) used domain experts as the method of verification. (Dukes et al., 2010) and (Muhammad, 2012) used Ibn Kathir as the method of verification.

Presently, most of the studies were carried out only to extract information from a single ontology. Most ontology matching systems are designed by assuming that the entities of both source and target ontologies are written in the same language (English, for instance). There is a lack of studies focused on matching from multi-lingual ontologies. With the increasing number of distributed resources, services, and applications on the web, multi-lingual ontology matching is likely to become important. Besides, most of the studies related to the Quran ontologies exist in Arabic and English. There is no existing ontology in Malay.

3. SYSTEM ACHITECTURE

3.1 Ontology-based Natural Language Processing for the Malay Quran:

Figure 1 below is a proposed system architecture for ontology-based natural language processing for the Malay Quran. The architecture contributes to better performance in retrieving relevant results from the Malay Quran. The design of this system architecture is adapted from the frameworks proposed by previous research (Alqahtani & Atwell, 2016; Uthayan & Mala, 2015;Yahya et al., 2013;Hamed & Aziz, 2016). This architecture is divided into two processes: (1) ontology construction and (2) query processing. Each processes has its own task.

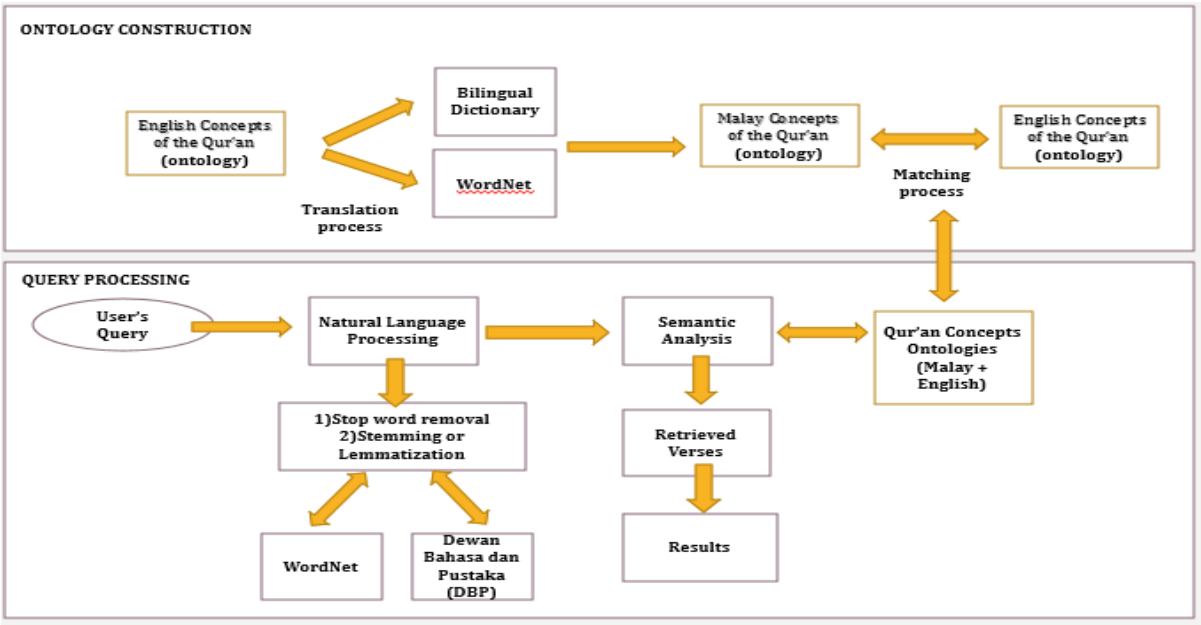

Figure 1: The proposed system architecture for ontology-based natural language processing for the Malay Quran

a) Ontology Construction Stage

In this phase, the existing Arabic and English Quran ontology translated into Malay. This is because the current ontology is only available in English and Arabic. In this process, the available concepts of the Quran will be used as resources, such as (Abbas, 2009,2013) , (Dukes, 2012,2013) and (Hakkoum & Raghay, 2015). The translation process will use available resources such as WordNet and Bilingual Dictionary. This process is important to get the right meaning for each concept. Next, the validation process will be conducted with Malay Quran experts to ensure the translated ontology terms are correct.

b) Query Processing Stage

A query will go through several Natural Language Processing (NLP) processes. Figure 2 below shows the processes that will be involved in query processing. Presently, there are no available or accessible NLP tools for Malay. So, the stop word removal, stemming and lemmatization will be constructed. Then, all the words will be analyze using a semantic analysis technique. This Semantic analysis technique deals with the meaning of words and sentences, the ways that words and sentences refer to elements in the world. In this phase, semantic analysis maps the analyzed words in the NLP to the concepts in Quran ontologies. At this time, it will search in Quran ontologies by using SPARQL to find the concepts related to the analyzed query and return the results.

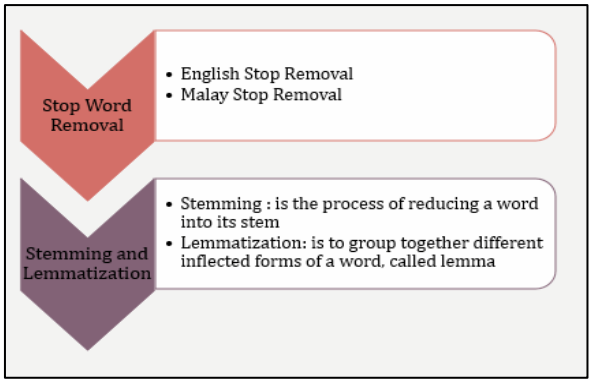

Figure 2: The processes in Natural Language Processing stage

The proposed system architecture will act as a guideline in conducting this research. It will assist in achieving the research aims and objectives. The architecture will present better processes that contributes to better performance in retrieving relevant results on Quran especially in Malay language.

3.2 Experiment: Searching for Synonyms in English and Malay Languages

The aim of this experiment is to observe word correlation in English and Malay translation of Quran. This observation is important to demonstrate one of the problems of existing keyword search techniques. For this experiment, we used the popular English translation of the Quran by Abdullah Yusuf Ali and Malay translation of the Quran from International Islamic University Malaysia (IIUM) as our data-set. For this initial stage, the translation of Surah Al-Baqarah, the largest chapter of Quran has been taken as a sample text. All the data stored and retrieved using Oracle 11g. In this experiment, we have chosen the concepts of Afterlife which is الجنة 'Jannah' and Hell. According WordReference.com, 'Jannah' means heaven or paradise and in Malay language, it is called as 'syurga'.

Firstly, to get the verses containing or relating to our target concept, two scholarly sources have been used; list of concepts from (Abbas, 2009, 2013) and the thematic index from *Uthmani* Malay Version of Quran (reference: text book). This thematic index has classified verses of Quran according to the themes in the Malay language whereby the list of concepts of (Abbas, 2009, 2013) has classified the verses according to Quran in Arabic and English. The process is carried out manually by checking every verses containing the concept from both sources. For instance, from the Malay thematic index, the word 'syurga' can be found in verse 5, 25, 38 etc. All the verses found in this process will be classified as relevant.

Secondly, the automated extraction from database using keyword search techniques was performed. The data in the database is derived from online sources which are popular English translation of the Quran by Abdullah Yusuf Ali and Malay translation of the Quran from International Islamic University Malaysia (IIUM). The aim of this automatic extraction is to get the verses from the online sources (classified as retrieved) and then compare it with the verses which have been classified as relevant. Table 2 shows the number of retrieved,

relevant and missed verses retrieved using keyword search for the concepts of 'Jannah' in English and Malay.

Table 2: Comparison between Relevant, Retrieved and Missed verses using keyword search.

Language	Keyword	Total No. of Relevant Verses	Retrieved	Missed
Malay	Syurga	13	10	3
Malay	Neraka	18	16	2
English	Heaven	10	11	1
English	Paradise	11	1	10
English	Hell	18	1	17

Based on the results on Table 2, there is a big difference between the number of retrieved and relevant verses in English words for Paradise and Hell. For word Paradise and Hell, only one verse had been retrieved by the system where the relevant was 11 and 18. Besides, for the word 'syurga', the difference between relevant and retrieved is 3. Subsequently, we analyze each relevant and retrieved verse. This analysis process is to find and compare the other word used by both English and Malay translation of Quran that represent the concepts of 'Jannah'. Results (see figure 3) show the other words used to describe 'syurga' in English terms are as Hereafter, The Garden, Paradise, and Home. In this case, The Garden is most widely used and has high similarities with 'syurga'. Surprisingly, there are no connection or relation between the word 'heaven' and 'syurga'. This is surprisingly unexpected because Malays normally used the word 'heaven' to describe 'syurga'. However, in this Malay Quran translation, the word 'heaven' has been translated as 'langit' (sky). So, it is not matched with the verses that contains the word 'syurga'. In this scenario, when the user use heaven as the keyword, they will get all the information about sky rather than information about 'syurga'. This occurs when the keyword search is the only technique used in search for information.

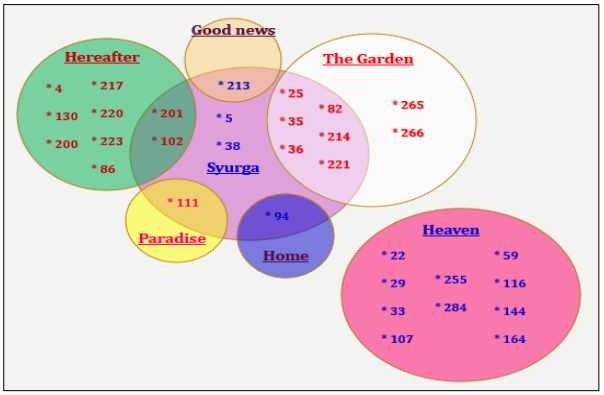

Figure 3: The other words used to describe 'syurga', heaven and paradise

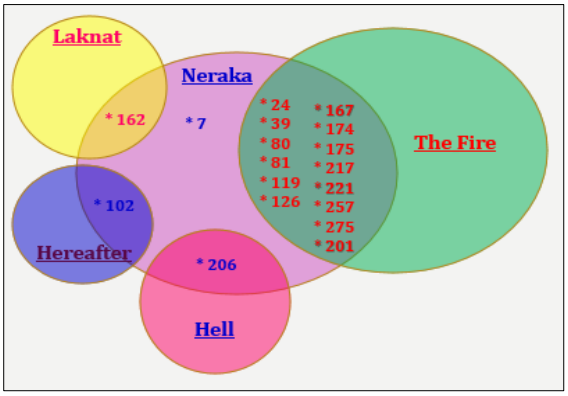

Figure 4: The other words used to describe 'neraka' and hell.

The other concepts of Afterlife is Hell. The word hell is often used to describe 'neraka' in Malay. However, the analysis results turned out differently (see figure 4). The most widely used word in this translation is 'The Fire'. Out of 18 verses, 14 used this word. In this case,

when a user submits the query using the word hell, he or she will get only one verse (verse 206).

As for the conclusion of this experiment, our key findings are:

(1) The translation plays a vital role in providing true and correct interpretation in every verse in the Quran. If not, it can cause confusion, especially the translation with two languages such as English and Malay. The widely spoken words must be checked regularly with the authorizing body whether it is right to be used and whether it is used in a correct meaning.

(2) The keyword based search alone is not enough to provide accurate and relevant results. It requires a combination with semantics and it takes from several sources that are authentic such as *Tafsir Ibn Kathir*.

(3) The lack of word translation equivalence between Malay and English is one of the problems derived from this experiment.

(4) The Malay retrieval score is higher than the English retrieval score. This indicating more use of metaphor in the English translation.

4. CONCLUSION & FUTURE WORK

This paper summarizes the existing problems related to the study of Quran Information Retrieval. Additionally, this paper also studied the current search techniques used in existing research on the Quran, especially the Malay Quran. This paper proposed new architecture for ontology-based natural language processing for the Malay Quran. With this architecture, it can help research on the Quran, mainly on the Malay Quran. This review helps the research in addressing the general problems and limitations in Quran Information Retrieval that influence its accessibility.

5. REFERENCES

Abbas, N., 2009. Qurany: A tool to search for concepts in the Quran. MSc Research Thesis, School of Computing, University of Leeds.

Abbas, N.H. & Atwell, E., 2013. Annotating the Arabic Quran with semantic web content tags. In E. Atwell & A. Hardie, eds. Proceedings of WACL-2 Second Workshop on Arabic Corpus Linguistics. Lancaster, UK, pp. 54–55.

Adhoni, Z. A., & Al Hamad, H. A. (2014). A Cloud Qur'an Application Using Drupal Technology. International Journal of Web Applications Volume, 6(1), 23–38.

Alfred, R., Mujat, A., & Obit, J. H. (2013). A Ruled-Based Part of Speech (RPOS) Tagger for Malay Text Articles. In Intelligent Information and Database Systems (pp. 50–59). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-36543-0_6

AlMaayah, M., Sawalha, M., & Abushariah, M. A. M. (2014). A Proposed Model for Quranic Arabic WordNet. In Proceedings of the 2nd Workshop on Language Resources and Evaluation for Religious Texts (pp. 9–13).

Alqahtani, M., & Atwell, E. (2015). A Review of Semantic Search Methods to Retrieve Information from the Qur ' an Corpus. In corpus Linguistics 2015 (pp. 7–9).

Alqahtani, M., & Atwell, E. (2016). Aligning and Merging Ontology in Al-Quran Domain. In 9th Saudi Students conference. Retrieved from http://eprints.whiterose.ac.uk/id/eprint/94921

Alrehaili, S. M., & Atwell, E. (2014). Computational ontologies for semantic tagging of the Quran : A survey of past approaches. In Ninth International conference on language resource and evaluation(LREC'14) (pp. 2–6).

Atwell, E., Habash, N., Louw, B., Abu Shawar, B., McEnery, T., Zaghouani, W., & El-Haj, M. (2010). Understanding the Al-Quran: A new Grand Challenge for Computer Science and Artificial Intelligence. In ACM-BCS Visions of Computer Science 2010. ACM-BCS Visions of Computer Science 2010.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). Modern Information Retrieval: The Concepts and Technology behind Search. In Information Retrieval (Vol. 82, p. 944). Pearson. http://doi.org/10.1080/14735789709366603

Dukes, K. (2013). Statistical Parsing by Machine Learning from a Classical Arabic Treebank. PhD Dissertation, School of Computing, University of Leeds. Retrieved from http://www.kaisdukes.com/papers/thesis-dukes2013.pdf

Dukes, K., Atwell, E., & Sharaf, A. B. M. (2010). Syntactic Annotation Guidelines for the Quranic Arabic Dependency Treebank. In Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt, 1822–1827.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2), 199–220. http://doi.org/10.1.1.101.7493

Hakkoum, A., & Raghay, S. (2015). Ontological approach for semantic modeling and querying the Qur'an. In 3 rd International Conference on Islamic Applications in Computer Science and Technologies (IMAN'15).

Hamed, S. K., & Aziz, M. J. A. (2016). A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification. Journal of Computer Science, 12(3), 169–177. http://doi.org/10.3844/jcssp.2016.169.177

Hargood, C., Millard, D. E., & Weal, M. J. (2008). A Thematic Approach to Emerging Narrative Structure. In Proceedings of the hypertext 2008 workshop on Collaboration and collective intelligence. http://doi.org/10.1145/1379157.1379168

Hasmy, H., Abu Bakar, Z., & Ahmad, F. (2015). Construction of Computational Lexicon for Malay Language. International Visual Informatics Conference, 257–268.

Ismail, R., Abu Bakar, Z., & Abd Rahman, N. (2015). Extracting knowledge from english translated Quran using NLP pattern. Jurnal Teknologi, 77(19), 67–73. http://doi.org/10.11113/jt.v77.6515

Manning, C. D., Prabhakar, R., & Schuatze, H. (2008). Introduction to Information Retrieval. Computational Linguistics (Vol. 35). http://doi.org/10.1162/coli.2009.35.2.307

M Tarawneh, & Al-Shawakfa, E. (2015). A Hybrid Approach for indexing and searching the Holy Quran. Jordanian Journal of Computer and Information Technology (JJCIT), 1(1), 41–50.

Muhammad, A. B. (2012). Annotation of conceptual co-reference and text Mining the Qur'an. PhD Dissertation, School of Computing, University of Leeds. Retrieved from http://etheses.whiterose.ac.uk/id/eprint/4160

Nassourou, M. (2011). A Knowledge-based Hybrid Statistical Classifier for Reconstructing the Chronology of the Quran. In WEBIST Special Session on Web and Text Mining (WTM 2011).

Nassourou, M. (2012). Using Machine Learning Algorithms for Categorizing Quranic Chapters by Major Phases of Prophet Mohammad's Messengership. International Journal of Information and Communication Technology Research, 2(11), 863–871.

Omar, N. (2011). Statistical Malay Part-of-Speech (POS) Tagger using Hidden Markov Approach. 2011 International Conference on Semantic Technology and Information,

231–236.

Othman, R., & Wahid, F. A. (2011). Issues in Evaluating the Retrieval Performance of Multiscript Translation of Al-Quran. 6th World Congress of Muslim Librarians and Information Scientist (WCOMLIS), i.

Saad, S., Salim, N., & Zainal, H. (2009). Pattern extraction for islamic concept. In Proceedings of the 2009 International Conference on Electrical Engineering and Informatics, ICEEI 2009 (Vol. 2, pp. 333–337). http://doi.org/10.1109/ICEEI.2009.5254719

Saad, S., Salim, N., Zainal, H., & Noah, S. A. M. (2010). A framework for Islamic knowledge via ontology representation. In 2010 International Conference on Information Retrieval & Knowledge Management (CAMP) (pp. 310–314). http://doi.org/10.1109/INFRKM.2010.5466897

Shoaib, M., Yasin, M. N., Hikmat Ullah, K., Saeed, M. I., & Khiyal, M. S. H. (2009). Relational WordNet model for semantic search in Holy Quran. 2009 International Conference on Emerging Technologies, ICET 2009, 29–34. http://doi.org/10.1109/ICET.2009.5353208

Ta'a, A., Zainal Abidin, S., Abdullah, M. S., Mat Ali, A. B., & Ahmad, M. (2012). Al-Quran Themes Classification Using Ontology. In 4th International Conference on Computing and Informatics (ICOCI 2013).

Tabrizi, A., & Mahmud, R. (2013). Coherence Analysis Issues on English-Translated Quran. Communications, Signal Processing, and Their Applications (ICCSPA), 1–6. Retrieved from http://umexpert.um.edu.my/file/publication/00001321_92836.pdf

Tian Tian. (2012). Using an ontology to improve the web search experience. PhD Dissertation, New Jersey Institute of Technology.

Ullah Khan, H., Muhammad Saqlain, S., Shoaib, M., & Sher, M. (2013). Ontology Based Semantic Search in Holy Quran. International Journal of Future Computer and Communication, 2(6), 570–575. http://doi.org/10.7763/IJFCC.2013.V2.229

Uthayan, K. R., & Mala, G. S. A. (2015). Hybrid Ontology for Semantic Information Retrieval Model Using Keyword Matching Indexing System. The Scientific World Journal, 2015(414910), 9.

Wahid, F. A. (2014). Retrieval Performance Of Quranic Texts (Pimpinan Ar-Rahman) In Jawi And Rumi Malay. PhD Dissertation, Kulliyyah of Information and Communication Technology

International Islamic University Malaysia.

Xian, B. C. M., Lubani, M., Ping, L. K., Bouzekri, K., Mahmud, R., & Lukose, D. (2016). Benchmarking Mi-POS: Malay Part-of-Speech Tagger. International Journal of Knowledge Engineering, 2(3), 115–121. http://doi.org/10.18178/ijke.2016.2.3.064

Yahya, Z., Abdullah, M. T., Azman, A., & Kadir, R. A. (2013). Query translation using concepts similarity based on Quran ontology for cross-language information retrieval. Journal of Computer Science, 9(7), 889–897. http://doi.org/10.3844/jcssp.2013.889.897

Yauri, A. R., Kadir, R. A., Azman, A., Azrifah, M., & Murad, A. (2012). Quranic-based Concepts : Verse Relations Extraction using Manchester OWL Syntax. In 2012 International Conference on Information Retrieval & Knowledge Management (pp. 317–321).

Yauri, A. R., Kadir, R. A., Azman, A., & Murad, M. A. A. (2014). Semantic Web Application for Historical Concepts Search in Al-Quran. International Journal on Islamic Applications in Computer Science And Technology, 2(2), 1–7.

Yunus, M. A., Zainuddin, R., & Abdullah, N. (2010). Semantic query with stemmer for quran documents results. ICOS 2010 - 2010 IEEE Conference on Open Systems, (Icos), 40–44. http://doi.org/10.1109/ICOS.2010.5720061