

# Exploring Meaningful Concepts of Al-Baqarah Chapter Using Text Mining Approach

Azman Ta'a and Shahrizal Sudin

Universiti Utara Malaysia, Malaysia, {azman@uum.edu.my, shahrizalum@gmail.com}

## ABSTRACT

This paper portrays the process of Natural language processing (NLP) to explore the meaningful concept in Al-Baqarah chapter of Al-Quran. Understanding the concept defined from the verses of Al-Quran will help users to learn Al-Quran content. Traditionally, the learner needs help from an Islamic Scholar to determine and understand the concept of Al-Quran verses. However, current tools lack to help user to extract and explore the meaningful concepts of the Al-Qur'an verses. This paper elaborates the using of text mining approach for identifying the meaningful concept such as prophets, messengers and fasting. Essentially, the analysis of meaningful concepts in Al-Baqarah chapter was performed using text mining approach in R Programming. The benchmarking approach has been used to validate the meaningful concepts produced from the analysis. The validation method compared the similarity of meaningful concept with the existing Al-Quran Malay translation text. The results have shown the similarity of meaningful concepts produced from text mining analysis with the existing Al-Quran textbooks.

**Keywords:** Natural language processing, text mining, meaningful concept, analytic programming, Al-Quran

## I INTRODUCTION

The Al-Quran is the holy book of Muslims, and Islamic scholars have described the Al-Quran as "It is the word of Allah SWT revealed to the prophet Muhammad SAW. His words are miracles, get rewarded when read, written in the books (Mushaf), narrated by connecting chain (Mutawatir)" (Mujahid, 2013). This holy book teaches moral, purification, good deeds, as well as those forbidden by the Almighty Allah. As Allah said in the following verses:

هَذَا بَصِيرَةٌ لِلنَّاسِ وَهُدًى وَرَحْمَةٌ لِّقَوْمٍ يُوقِنُونَ

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

(Al-Quran - Al-Jaathiyah 45:20)

The Al-Quran provides guidance to mankind, promotes justice between one another, and provides guidance on how to live on earth and with neighbours. A related study described the Al-Quran as a source of information on any subject matter concerning the world and the hereafter (Atwell et al., 2010; Ta'a, Abed, & Ahmad, 2017). However, knowledge in the Al-Quran cannot be compared with scientific books because the former provides real and deep discussions of matters under the examination of the almighty God Allah SWT. Therefore, an acquiring of the meaningful concepts in Al-Quran verses is important for reader to learn and understand the Al-Quran knowledge.

Currently, the searching of meaningful concept in Al-Quran using a keyword is either unclear or inaccurate (Khan, Saqlain, Shoaib, & Sher, 2013). The searching was not based on concept classification and clustering, and these aspects are the major shortcomings of existing issues related to the retrieval of knowledge from the Al-Quran. Several researches had used ontology and semantic approach to perform the knowledge retrieval of Al-Quran (Saad, Salim, Zainal, & Muda, 2011; Ta'a, Abed, & Ahmad, 2017). However, the results are still unconvincing, and this requires a proper extraction of meaningful concepts from the Al-Quran verses prior the searching can be performed.

Several research has studied the Arabic text of Quran and deal with the translation of the meaning of the Al-Quran verses. Dukes, Atwell, & Habash (2013) has developed an open source Quranic corpus using both Arabic as well as translation of these words. The Quranic Arabic corpus consist of part-of-speech tagging, morphological segmentation and syntactic analysis using reliance grammar, that enables produced a source further analysis of the Quran. For a Muslim, lot of problems related to religion are difficult to solve due to misunderstanding of the Al-Qur'an contents. Moreover, reading the Al-Qur'an in a conventional way and wrong interpretation, especially without the help from Islamic scholars. For example, a paragraph contain the verses in the Al-Qur'an are dispersed, so it takes a long time to search for required information. The process of searching the verses in the Al-Qur'an in a conventional way or with digital Al-Qur'an, which have been available on the internet do not help much, as the given

information is not relevant to the problem faced. Therefore, it needs a system for identifying, searching and classifying the concept written in the Al-Qur'an verses.

The current scenario of learning Al-Quran is using traditional methods, which is the learning process is face to face with the teachers. The concept about the verses was manually identified and verbally explained by the teachers. However, it will be easier for learner if they can identify the meaningful concept by using tools. Since, the Al-Qur'an was translated into several languages by Muslim experts, therefore, by using tools that support the process to identify the concept for the verses (e.g., prayer, fasting, law, and hajj), the learner can enhance their learning process and help to understand Al-Quran in a flexible way. Even current tools already have searching facilities for Al-Quran words, however this do not help much when the searching results was not corresponded to the learner requirements. For this reason, this project will using a text analysis process called text mining, which consist of several methods such as parsing, stemming, and morphing to produce the meaningful words. Ultimately, the text mining process will identify the concepts or terms pertaining to the verses. There are several issues related to the Malay translation of the Al-Qur'an, which is about the standardization version of translations itself. This scenario provides the different structures of meanings due to the translation was made by different Islamic scholars. However, this paper has selected the recent Malay translation textbook of Al-Quran.

## II RELATED WORK

This paper discussed the current works of the Al-Qur'an text analysis of Malay translation, review on the previous work, and explore the basic techniques of Natural Language Processing (NLP) such as text processing, stop words, stemming algorithm, and ngram (Bakhtin, Szlam, Ranzato, & Grave, 2018). This will help to extract and analyze the meaningful concept contained in the Malay translation of Al-Qur'an. The Al-Qur'an has been translated into various languages around the world by the Muslim scholar. Particularly in Malaysia, the readers are required to read and learn the meaning of Al-Qur'an in Malay translation language. Even though there are various version translation of the Al-Qur'an, they do not help the reader to understand because some other version was presented the concept in different meaning. Few issues such as ambiguities of word, lack of word equivalent to other language such as between Malay and Arabic or Malay and English, different structure of word, sentences and discourse in these languages.

The paper aims to provide the helpful and accurate meaningful concept of Al-Quran using a framework in the field of Arabic Natural Processing (ANP). The researcher try to find approach and experiment for analyzing the Arabic text and then provide statistic information by using several technique in NPL such as text preprocessing and text mining operations. However, in current NLP process, current tools does not support stop word and stemming functions for Arabic language (Adhoni, Al Hamad, Siddiqi, Parvez & Adhoni, 2013; Alhawarat, Hegazi, & Hilal, 2015). Thus, the simple cleaning has been applied on the Al-Quran corpus and normalizing some of the Arabic words. The result to reveal simple fact about the term of Arabic Qur'an has shown a diversity features of the Al-Qur'an, such as an important words with highest term frequencies using wordcloud utility. The word frequency have been calculated using both Term Frequency (TF) and Term Frequency-inverse Document Frequency (TF-IDF) methods, which was supported most of the current text analytic tools.

## III EXTRACTING MEANINGFUL CONCEPT

The process to determine meaningful concept of Al-Quran is based on the Natural Language Processing (NLP) approach. The meaningful concept is defined from text translation of Al-Quran that explained the content or message of verses (Khajehei & Shakarami, 2012). Most current Al-Quran textbooks has provided translation and concepts for the block of verses. The NLP is a method to process and organize the Al-Qur'an text translation to produce the meaningful concepts of Al-Baqarah chapter. Text analysis has adapted the ngram technique to produce the frequency list of words of Al-Quran text to determine the meaningful concept of verses. The whole process to extract and analyze the Al-Quran text is discussed in the following detail.

### A. Preparing Al-Quran Corpus Text

In the first stage, the Al-Qur'an Malay text translation is read and converted to a corpus format and perform a text preprocessing. The text preprocessing aims to provide clean text data, thus later it can produce a useful information from text documents (i.e., meaningful concepts). The idea in the second stage is to implement and evaluate the Al-Quran text translation as illustrated in the following model as shown in Figure 1.

### B. Malay Al-Quran Translation Text: Al-Baqarah Chapter

Al-Baqarah chapter is a Madani (Madinah) verses and the largest chapter in Al-Quran and contains a wide range of topic from legislations to stories of

Prophet Muhammad. Some information about the Al-Baqarah chapter is stated in Table 1.

For the text sampling, the input sentences have taken from The Al-Karim Amazing 33 in 1 Al-Qur'an, which has been certified by the Department of Islamic Development Malaysia (JAKIM) (Mustapha, 2013). The ideal solution for scanning Al-Qur'an Amazing is by using Optical Character Recognition (OCR), which is essentially the conversion of scanned images with text that allows to extract data from Amazing Al-Qur'an that scanned documents and the text is saved to be used for text analytics process. After finishing the NPL process, the produced text was evaluated using a benchmarking approach, which is the produced text (referred as term or concept) will be compared with the concept defined by the selected Al-Quran Malay translation textbook.

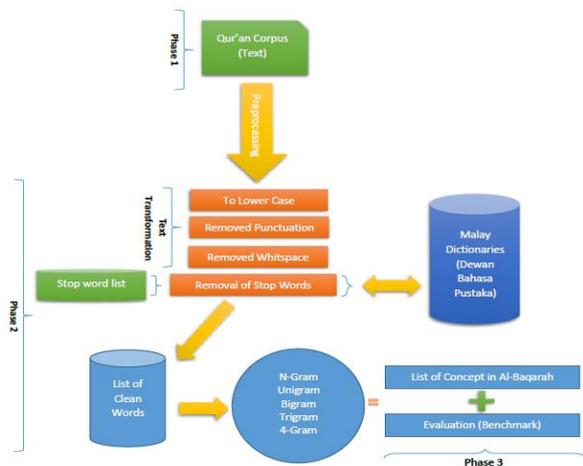


Figure 1. Extracting Meaningful Concept Model.

Table 1. Information about Al-Baqarah Chapter

No. of verses	286
No. of words	6141
Stories mentioned	Story of creation, Moses(Musa), Abraham (Ibrahim), Israel, Saul and Goliath
Rulings mentioned	Prayer, marriage and divorce, financial transactions, fasting and haj (pilgrimage)

The main structure for managing text document in R is so-called Corpus that utilizing framework provided by R package called tm. A Corpus is an abstract concept, representing a collection of text document. The other argument is read lines of the corpus constructor to read source of text as shown in Figure 2.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
ngram-malaystopword.R
29 alquran <- readLines("SurahAlBaqarah.txt")
30
31 albaqarah <- corpus(vectorSource(alquran))
32
33 #Cutom Malay stopwords
34 malaystopwords <- readLines("StopwordID.txt")
35 class(malaystopwords)
36 malaystopwords_vector <- as.vector(malaystopwords)
37
38 albaqarah <- tm_map(albaqarah, content_transformer(tolower))
39 albaqarah <- tm_map(albaqarah, removePunctuation)
40 albaqarah <- tm_map(albaqarah, removeNumbers)
41 albaqarah <- tm_map(albaqarah, stripWhitespace)
42 albaqarah <- tm_map(albaqarah, removeWords, c(malaystopwords_vector, stopwords("english")))
43
  
```

Figure 2. Text Analysis in R

Text preprocessing transformations are using these techniques: *tolower*, *removePunctuation*, *removeNumbers*, *stripWhitespace*, and *removeWords* (Malay stop word). The stop word is a list of the highest frequency function words assumed to have no indexing value. Furthermore, the 36 stop words that were added in the list (original) containing 314 stop words, which totaling 350 stop words. Table 1 show examples of the Malay stop word, which most of the words have similar meaning (Taufik, Abdullah, & Fatimah, 2005; Yahaya, Rahman, & Bakar, 2017).

Table 1. Example Malay stop word

ada	Inikah	Sampai	adakah
adalah	Itukah	Sangatlah	adanya

### C. N-Gram Frequency

N-gram analysis is used to determine the most frequent words in the Al-Quran starting from single word (unigram), two words (bigram), 3 words (trigram), four words (fourgram), and so on. The most frequency words attempts to calculate the term frequency in Al-Baqarah chapter by using Term Frequency (TF) method. The calculation used to measure the term document matrix (Alhawrat, Hegazi & Hilal, 2015). Example of the R codes for ngram functions is shown in Figure 3.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
ngram-malaystopword.R
65 BigramTokenizer <- NgramTokenizer(albaqarah, weka_control(min = 2, max = 2))
66 two_word <- data.frame(table(BigramTokenizer))
67 sort_two <- two_word[order(two_word$Freq, decreasing = TRUE),]
68 wordcloud2(sort_two, size=0.3, color="random-dark")
69 wordcloud2(sort_two, size = 0.7, shape = "star")
70 # change the background color
71 wordcloud2(sort_two, size=1.6, color="random-light", backgroundColor="black")
72 # change the rotation wordcloud
73 wordcloud2(sort_two, size = 0.5, minRotation = -pi/2, maxRotation = pi/2)
74 wordcloud2(sort_two, size = 0.5, minRotation = -pi/6, maxRotation = pi/6, rotateRatio=1)
75 head(two_word, 20)
76 head(sort_two, 100)
77
78 TrigramTokenizer <- NgramTokenizer(albaqarah, weka_control(min = 3, max = 3))
79 three_word <- data.frame(table(TrigramTokenizer))
80 sort_three <- three_word[order(three_word$Freq, decreasing = TRUE),]
81 wordcloud2(sort_three, size=0.3, color="random-dark")
82 wordcloud2(sort_three, size = 0.7, shape = "star")
83 # change the background color
84 wordcloud2(sort_three, size=1.6, color="random-light", backgroundColor="black")
85 # change the rotation wordcloud
86 wordcloud2(sort_three, size = 0.2, minRotation = -pi/2, maxRotation = -pi/2)
87 wordcloud2(sort_three, size = 0.2, minRotation = -pi/6, maxRotation = -pi/6, rotateRatio=1)
88 head(three_word, 50)
89 head(sort_three, 100)
90
91 fourgramTokenizer <- NgramTokenizer(albaqarah, weka_control(min = 4, max = 4))
92 four_word <- data.frame(table(FourgramTokenizer))
93 sort_four <- four_word[order(four_word$Freq, decreasing = TRUE),]
  
```

Figure 3. Implement N-Gram in R

The experiment of unigram, bigram, trigram and fourgram words are visualized using wordcloud function in R. Wordcloud is capable to visualize the



Then, the benchmarking is performed according to the following algorithm:

- Step 1: Start
- Step 2: Read Al-Furqan/Al-Humaira
- Step 3: Read Unigram result
- Step 4: Perform a comparison between Unigram and Al-Furqan/Al-Humaira using pattern matching function
- Step 5: Perform an analysis to find the matching of concepts in unigram and Al-Furqan/Al-Humaira
- Step 6: End

The result produced from these benchmarks are shown in Figure 5, Figure 6, Figure 7, and Figure 8.

	allah	orang-orang	muhi	orang	orang-orang	beriman	lailan	balasan	berang	engapa	muhammad	teguh	maris
golongan orang muslim...	0.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
golongan orang muslim...	0.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah menyuruh Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah menyuruh Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
lailan Allah s.w.t terhadap kaum muslimin menges...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
gagasan terhadap orang yg beriman...	0.0	0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perumpamaan dan hikmah...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
hikmah Allah s.w.t terhadap Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah manusia dan Tuhan terhadap Khalifah...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah dan larangan Allah s.w.t kepada bani israil...	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
nikmat Allah s.w.t kepada bani israil...	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
balasan terhadap perbuatan dan sikap bani israil...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
balasan terhadap perbuatan dan sikap bani israil...	0.0	0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
balasan terhadap bani israil karena melanggar perintah...	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
buah penyembelitan lembu betina...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
kegiatan terhadap manusia orang jahili...	0.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
bani israil melanggar perintah dengan Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
halq orang jahili terhadap para nabi dan Malaikah...	1.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

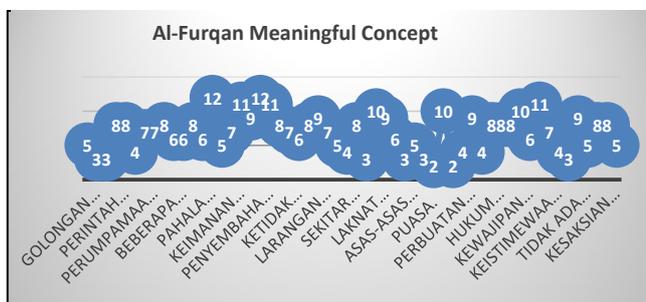


Figure 5. Matching meaningful concept in Al-Furqan

	allah	orang-orang	muhi	orang	orang-orang	beriman	lailan	balasan	berang	engapa	muhammad	teguh	maris
golongan orang muslim...	0.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
golongan orang muslim...	0.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah menyuruh Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah menyuruh Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
lailan Allah s.w.t terhadap kaum muslimin menges...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
gagasan terhadap orang yg beriman...	0.0	0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perumpamaan dan hikmah...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
hikmah Allah s.w.t terhadap Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah manusia dan Tuhan terhadap Khalifah...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
perintah dan larangan Allah s.w.t kepada bani israil...	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
nikmat Allah s.w.t kepada bani israil...	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
balasan terhadap perbuatan dan sikap bani israil...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
balasan terhadap perbuatan dan sikap bani israil...	0.0	0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
balasan terhadap bani israil karena melanggar perintah...	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
buah penyembelitan lembu betina...	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
kegiatan terhadap manusia orang jahili...	0.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
bani israil melanggar perintah dengan Allah s.w.t.	1.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
halq orang jahili terhadap para nabi dan Malaikah...	1.0	0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

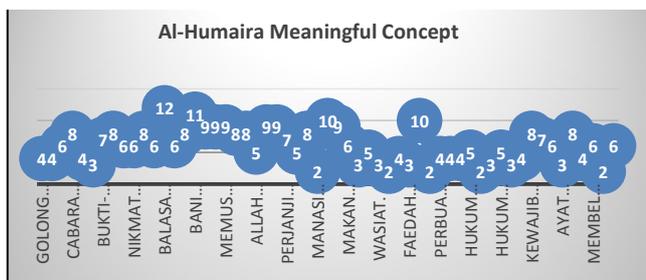


Figure 6. Matching meaningful concept in Al-Humaira

Based on the benchmarks can determine most of word produced from the ngram technique has produced the meaningful concept from the Al-Quran (i.e., Al-Baqarah chapter). The matching similarity of unigram with concept in Al-Furqan and Al-Humaira are calculated using these formula:

1. Matching Similarity (%) = (Total of unigram match in Al-Furqan / Total of unigram result) X 100
2. Matching Similarity (%) = (Total of unigram match in Al-Humaira / Total of unigram result) X 100

Based on the formula the matching similarity of benchmarks are calculated as follow:

<b>The benchmarking is prepared with the following datasets:</b>
1. Total of unigram match in Al-Furqan (57 Concept) = <b>336</b>
2. Total of unigram match in Al-Humaira (56 Concept) = <b>378</b>
3. Total concept of unigram = 336+378 = <b>714</b> concepts

<b>Percentage for Al-Furqan</b> (336 / 714) X 100 = <b>47%</b>
<b>Percentage for Al-Humaira</b> (378 / 714) X 100 = <b>53%</b>

The results of benchmark 1 and benchmark 2 show that the matching concept between unigram and Al-Furqan and Al-Humaira was 47% and 53% each. The results are not high due to the unnecessary word still need to be excluded. This required further analysis for the unigram results in order to clearly produce the meaningful concepts. However, the pattern match do not guarantees the similarity of semantic concept between the concepts matching. The similarity of the semantic concept was not covered in this paper.

## VI CONCLUSION

This project has successfully developed the application for finding meaningful concepts in the Al-Quran (Al-Baqarah chapter) using the NLP approach. The ngram technique has produced the meaningful terms or concept, which is showing the most frequent words exist in the Al-Quran. The NLP process shows that all words produced by unigram have matched with a different frequency concept provided by Al-Furqan and Humaira translation. The methodology for achieving the objective of this project was implemented properly started with define the meaningful concept of Al-Qur'an definition, performed the NPL processing such as tokenization, cleaning the text, remove stop words and performed benchmarking for evaluating the similarity of concept existing in Malay translation Al-Quran textbook. Indeed, the application for finding a meaningful concept of Al-Quran will help learner to find the concept of any Al-Quran verses, and later be used for learning the Al-Quran contents. Future works will explore the using of semantic approach to find a more meaningful concept in the Al-Quran.

## REFERENCES

- Alhawarat, M., Hegazi, M., & Hilal, A. (2015). Processing the Text of the Holy Quran: a Text Mining Study. *International Journal of Advanced Computer Science and Applications*, 6(2), 2–7.
- Adhoni, Z. A., Al Hamad, H., Siddiqi, A. A., Parvez, M., & Adhoni, Z. A. (2013). Cloud-based Online Portal and Mobile Friendly Application for the Holy Quran. *Life Science Journal*, 10(12s).
- Atwell, E., Habash, N., Louw, B., Abu Shawar, B., McEnery, T., Zaghouni, W., & El-Haj, M. (2010). Understanding the quran: A new grand challenge for computer science and artificial intelligence. *ACM-BCS Visions of Computer Science* 2010.
- Ahmad, F. D. (1995). A Malay language document retrieval system: An experimental approach and analysis. PHD. Bangi: Universiti Kebangsaan Malaysia, pp 322.
- Bakhtin, A., Szlam, A., Ranzato, M. A., & Grave, E. (2018). Lightweight Adaptive Mixture of Neural and N-gram Language Models. arXiv preprint arXiv:1804.07705.
- Biru, T. (2015). *Al-Quran Al-Karim dan Terjemahan Al-FURQAN*. Kuala Lumpur: Telaga Biru Sdn. Bhd.
- Dukes, K., Atwell, E., & Habash, N. (2013). Supervised collaboration for syntactic annotation of Quranic Arabic. *Language Resources and Evaluation*, 47(1), 33–62. <https://doi.org/10.1007/s10579-011-9167-7>
- Enterprise, H. B. (2012). *Al-Quran Tajwid dan Terjemahan HUMAIRA*. Kajang, Selangor: Enterprise, Humaira BookStore.
- Herdianto, A. (2010). Pencarian Ayat-Ayat Alquran Berdasarkan Konten Menggunakan Text Mining Berbasis Aplikasi Desktop. *Pencarian Ayat-Ayat Alquran Berdasarkan Konten Menggunakan Text Mining Berbasis Aplikasi Desktop*, 2, 1–14.
- Ismail, N. K., Abdul Rahman, N., Abu Bakar, Z., & Tengku Sembok, T. M. (2007). Terms Visualization for Malay Translated Quran Documents. *Proceedings of the International Conference on Electrical Engineering and Informatics*, (January 2014), 554–557.
- Khajehei, H., & Shakarami, A. (2012). An In-Depth Linguistic Analysis of the Thematic Connectivity in the Holy Quran. *Academic Research International*, 2(2), 682–695. Retrieved from [http://www.savap.org.pk/journals/ARInt./Vol.2\(2\)/2012\(2.2-75\).pdf](http://www.savap.org.pk/journals/ARInt./Vol.2(2)/2012(2.2-75).pdf)
- Kassim, M. N., Maarof, M. A., Zainal, A., & Wahab, A. A. (2016). Word stemming challenges in Malay texts: A literature review. In *Information and Communication Technology (ICoICT), 2016 4th International Conference on* (pp. 1-6). IEEE.
- Ta'a, A., Abed, Q. A., & Ahmad, M. (2017). AL-QURAN ONTOLOGY BASED ON KNOWLEDGE THEMES. *Journal of Fundamental and Applied Sciences*, 9(55), 800–817.
- 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.
- Yahaya, F., Rahman, N. A., & Bakar, Z. A. (2017). Resolving Malay Word Sense Disambiguation Utilizing Cross-Language Learning Sources Approach. *Advanced Science Letters*, 23(11), 11320-11324.