EXTRACTING MEANINGFUL CONCEPT OF AL-QURAN USING TEXT MINING APPROACH: A CASE STUDY ON AL-BAQARAH CHAPTER

Shahrizal Sudin[1,a], Azman Ta'a[2,b]

[1,2]School of Computing, College of Arts and Sciences
Universiti Utara Malaysia (UUM)
06010 UUM Sintok, Kedah Darul Aman

[a]shahrizaluum@gmail.com, [b]azman@uum.edu.my

Keywords: Al-Quran, Text Analytic, Meaningful Concept, Malay Natural Language Processing, Analytic Programming

Abstract. This project portrays the process of Natural language processing (NLP) of text mining using analytic programming language to extract the term or concept in Al-Quran. Understanding the concept defined from the verses of Al-Quran will help users to learn Al-Quran in more faster. The current Al-Qur'an learning approach is using conventional method that needs help from an Islamic Scholar **'ULAMA'** to facilitate user to understand the concept of verses in the Al-Qur'an in easy manner. Obviously, the scholar relies on his personal knowledge and familiarity of the Qur'an contents. However, there is a lack of computational tools to help the reader to capture and understand the concepts inside the verse of the Al-Qur'an. Therefore, using text mining approach, it can support the process of identifying the meaningful concept behind the verses (e.g., Prophets, Messengers and fasting). This project elaborate the details of the task in different perspective by looking at the problem statement, objectives, research methodology, expected outcomes and significance of the output. Essentially, the extracting and evaluation of text analysis approach has been done by using powerful analytic programming language called R. The benchmarking method has been used to validate the result of this project that based on similarity of concept produced from the NLP process with the concept determined by the existing Al-Quran textbook published from two different publishers. The NLP process that based on unigram technique has successful produced the meaningful concepts of Al-Baqarah chapter, which is has the similarity with the concepts determined by the Al-Quran textbooks.

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:

هَٰذَا بَصَٰٓئِرُ لِلنَّاسِ وَهُدًى وَرَحْمَةٌ لِّقَوْمٍ يُوقِنُونَ ﴿٢٠﴾

> *This [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, Abdullah, Ali & Ahmad, 2014). 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 search and retrieval 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, Abdullah, Ali & Ahmad, 2014). However, the results are still unconvincing, and this requires a proper treatment of concepts (i.e., meaningful concepts) that to be extracted from the Al-Quran verses prior the semantic search can be performed.

Several research studies have studied the Arabic text of Quran, in fact many studies deal with the translation of the meaning of the words of the Quran. Kais Dukes and his team showed an open source Quranic corpus using both Arabic as well as translation of these words. The Quranic Arabic Corpus is a jointly built linguistic resource that began at Leeds University, with diversity layer of annotation consist of part-of-speech tagging, morphological segmentation and syntactic analysis using reliance grammar, that enables produced a source further analysis of the Quran (Dukes, Atwell, & Habash, 2013). For a Muslim a 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 Qur'an are dispersed, so it takes a long time to search for required information. The process of searching the verses in the Qur'an in a conventional way or with digital 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, 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 was provided the different structures of meanings due to the translation was made by different Islamic scholars. However, for this project we have selected the recent Malay translation textbook of Al-Quran.

RELATED WORK

This study discussed the current works of the al-Qur'an text analysis of Malay translation, to review on the previous work, and explore the basic techniques of Natural Language Processing (NLP) and text processing, stemming algorithm, and ngram techniques. This will help the challenge to extracting and analyzing the meaningful concept contained in the Malay translation of Al-Qur'an. The Holy Qur'an has been translated into various languages around the world by the Muslim scholar. In particular in Malaysia country as for Malays readers are required to read and learn the meaning of Al-Qur'an Translation Malay Language to understand the reference word of God (Allah). Even though there are various version translation of the meaning of the Al-Qur'an, they do not help the reader but some other version (concept) present different from them. However, there few issues such as ambiguities of word lack of word equivalent between Malay and Arabic or Malay with English and different structure of word, sentences and discourse in these two languages.

The research aims to provide the helpful and accurate meaningful information and knowledge of Al-Quran using framework in the field Arabic natural processing. The researcher try to find approach and experiment for analyzing Arabic text and then providing statically information by using several technique in NPL such as text preprocessing and text mining operations. In the NLP process, this study uses tools but there are also constraints in the process of NPL due to current tools does not support stop word and stemming for Arabic language (Muhammad Alhawarat, Anwer 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 by showing a diversity features of the Holy Al-Qur'an, such 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 text analytic tool (Muhammad Alhawarat, Anwer Hilal, 2015).

EXTRACTING MEANINGFUL CONCEPT OF AL-QURAN

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 (Setyarahajoe and Shakarami, 2012). Most current Al-Quran textbooks has provided translation and concepts for the verses. The NLP is a process to process and organize the Al-Qur'an text (the Malay translation) to produce the meaningful concepts of Al-Quran, especially in Al-Baqarah Chapter. The NLP process adapted the ngram technique to produce the frequency list of words of Al-Quran text to determine the meaningful concept of verses.

Preparing Al-Quran Corpus Text. The Al-Qur'an Malay text is read as a corpus and perform text preprocessing namely the process of its predecessor applied to the text that aims to produce useful information from text documents. The idea

in the second stage is to implement and evaluate Al-Quran text translation as illustrated in the following step Figure 1.

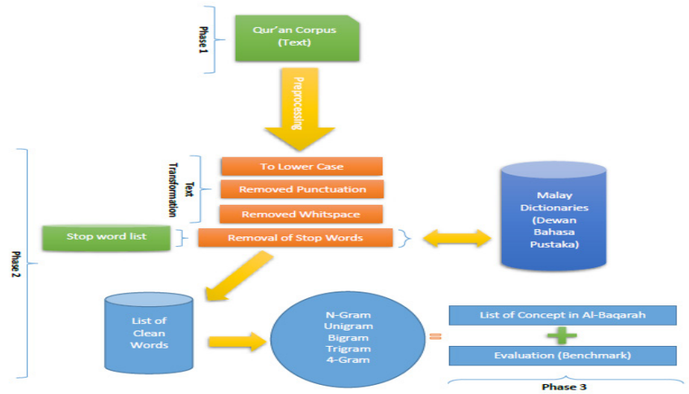

Figure 1: Implementation and evaluations model for Quran Application

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. Information about the Al-Baqarah chapter is described in Table 1.

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)

For the text sampling, the input sentences have been 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.

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

Figure 2: Text transformation process in R Programming

Text preprocessed ***transformations*** are using these techniques: tolower, removePunctuation, removeNumbers, stripWhitespace, and RemoveWords (Malay Stop Word). The stop word list is a list of the highest frequency function words assumed to have no indexing value. Furthermore, of the 36 stop words that were added in the list (original) containing 314 stop words, totaling 350 stop words. These are poor differential and is not possible to identify the contents of the document. Table 1 show examples of the Malay stop word (Muhammad Taufik Abdullah, Fatimah Ahmad, 2005).

Table 1: Example Malay stop word

ada	inikah	Sampai	adakah	inilah	sana
adalah	itukah	Sangatlah	adanya	itulah	saya

N-Gram Frequency. ***R-Weka*** package is used for n-gram to analysis 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 to measure the term document matrix (Alhawrat, Hegazi & Hilal, 2015). Example of the R codes for ngram functions is shown in Figure 3.

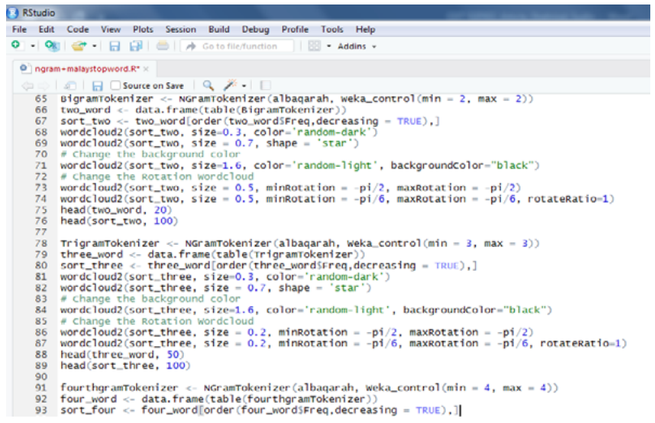

Figure 3: R codes for ngram

The experiment of unigram, bi-gram, tri-gram and four-gram words are visualized using wordcloud. Wordcloud capable to visualize the frequency of different words by colors and size of the words.

RESULT AND ANALYSIS

Different set of experiment were carried out on the Malay text of the Al-Baqarah Chapter. These experiments determine the frequency of words exist in the Al-Baqarah chapter based on the **term frequency** algorithm built in the ngram technique. The experiment manipulated the text of Al-Baqarah Chapter in order to produce list of words with the most frequent words presented in the Al-Baqarah chapter. Then, the frequency words are visualized in wordcloud with variation of frequency level. The wordcloud visualization is popular method to present the unigram, bigram, trigram and fourgram words in text analytics (Ismail & Rahman A, 2007). For example, the most frequency words of unigram in Al-Baqarah chapter is shown in Table 2.

Table 2. *The 100 most frequent words of Unigram in Al-Baqarah Chapter*

UnigramTokenizer	**Freq**						
21	allah	307	23	alquran	14	836	mendengar	9
1074	orangorang	159	344	diturunkan	14	872	mengambil	9
1073	orang	91	715	mati	14	912	mengikuti	9
138	beriman	71	1453	tuhannya	14	1075	pahala	9
148	berkata	51	606	ketahuilah	13	1258	saksi	9
1451	tuhan	42	1346	solat	13	1397	taurat	9
80	barang	37	2	adakah	12	1471	umat	9
1034	muhammad	36	112	berat	12	1485	wajah	9
457	ingatlah	33	191	bertakwalah	12	43	arah	8
699	manusia	32	251	dalamnya	12	77	bani	8
1197	petunjuk	29	356	dunia	12	121	berdoa	8
504	kafir	28	659	Ibrahim	12	468	isteri	8
629	kitab	26	1139	penyayang	12	484	janji	8
125	berfirman	25	1495	yahudi	12	549	kekal	8
228	bumi	23	6	agama	11	588	kerjakan	8
522	kebaikan	23	30	anak	11	605	ketahui	8
1436	tiada	23	414	harta	11	804	memperoleh	8
421	hati	21	573	kepadaku	11	867	mengajarkan	8
514	katakanlah	21	664	lembu	11	902	menghidupkan	8
350	dosa	19	939	menjadikan	11			
57	azab	18	1233	rezeki	11			
			1372	syurga	11			

Based on the Table 2, the most frequency words of unigram can be visualized using wordcloud as shown in Figure 4. Clearly show that "ALLAH" was the most frequent word in Al-Baqarah chapter. We also performed the bigram, trigram and fourgram analysis for comparing with unigram.

Figure 4: *Wordcloud for most frequent words in Al-Baqarah Chapter Using Unigram*

The above results solely based on counting analysis of the most frequent words in the Al-Baqarah chapter. Based on the results, we can conclude the word "Allah" be the top list in the ngram frequency emphasizing the center of Islam surround the monotheism creed ("Tauhid") because most words in bigram and trigrams have focused on the unity and glory of God. Although the result was produced some of the low ngram frequency value, this does not mean that the produced words was not indicated these words are meaningless. These lowest frequency words such as *Kiblat, Prayer, Fasting* and *Hajj* were the most important concepts that become the obligation to the Muslim (i.e., syariah rules). Meanwhile, the fourgram is the most difficult to determine the meaningful concepts due to ambiguity meaning of the word combination.

VALIDATION (BENCHMARKING)

The important process to find the similarity of meaningful word with the concepts determined in other Al-Quran textbook. This process has been performed by using ***grepl*** *Function* in R. The **grepl** functions returns ***TRUE (Value 1)*** if a string contains the pattern, otherwise it will return ***FALSE (Value 0)***, if the parameter is a string vector, returns a logical vector (match or not for each element of the vector). Therefore, this validation process performed the benchmarking method for unigram results only, since the bigram, trigram and fourgram results was not considered due to ambiguities meaning of words produced by these techniques. The unigram results is compared with the concept determined in two Al-Quran

translation textbook named Al-Furqan (Furqan, 2015) and Al-Humaira (Humaira, 2012). Then, the benchmark process identified the pattern matching between Unigram results *(single word as shown in Table 3*) with the Al-Furqan and Al-Humairah. The benchmarking is prepared with the following datasets:

1. Unigram result, which consist 1,467 words.
2. Al-Furqan, which consist of 57 concepts.
3. Al-Humairah, which consist of 56 concepts.

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	orangorang	maha	orang	sesungguhnya	beriman	tuhan	bahawa	barang	engkau	muhammad	ingatlah	manusia
golongan orang mukmin.	0	0	0	1	0	0	0	0	0	0	0	0	0
golongan orang munafik.	0	0	0	1	0	0	0	0	0	0	0	0	0
perintah menyembah allah s.w.t.	1	0	0	0	0	0	0	0	0	0	0	0	0
cabaran allah s.w.t terhadap kaum musyrikin mengenai ...	1	0	0	0	0	0	0	0	0	0	0	0	0
ganjaran terhadap orang yg beriman.	0	0	0	1	0	1	0	0	0	0	0	0	0
perumpamaan dan hikmah	0	0	0	0	0	0	0	0	0	0	0	0	0
bukti-bukti kekuasaan allah s.w.t.	1	0	0	0	0	0	0	0	0	0	0	0	0
penciptaan manusia dan tugas sebagai khalifah.	0	0	0	0	0	0	0	0	0	0	0	0	1
perintah dan larangan allah s.w.t kepada bani israel.	1	0	0	0	0	0	0	0	0	0	0	0	0
nikmat allah s.w.t kepada bani israel.	1	0	0	0	0	0	0	0	0	0	0	0	0
balasan terhadap perbuatan dan sikap bani israel.	0	0	0	0	0	0	0	0	0	0	0	0	0
pahala bagi orang yang beriman.	0	0	0	1	0	1	0	0	0	0	0	0	0
balasan terhadap bani israel kerana melanggar perjanji...	1	0	0	0	0	0	0	0	0	0	0	0	0
kisah penyembelihan lembu betina.	0	0	0	0	0	0	0	0	0	0	0	0	0
keraguan terhadap keimanan orang yahudi.	0	0	0	1	0	0	0	0	0	0	0	0	0
bani israel mengingkari janjinya dengan allah s.w.t.	1	0	0	0	0	0	0	0	0	0	0	0	0
sikap orang yahudi terhadao para rasul dan kitab yang ...	1	0	0	1	0	0	0	0	0	0	0	0	0

Figure 5: Comparison between Unigram and Al-Furqan

	allah	orangorang	maha	orang	sesungguhnya	beriman	tuhan	bahawa	barang	engkau	muhammad	ingatlah	manusia
golongan orang mukmin.	0	0	0	1	0	0	0	0	0	0	0	0	0
golongan orang munafik.	0	0	0	1	0	0	0	0	0	0	0	0	0
perintah menyembah allah s.w.t.	1	0	0	0	0	0	0	0	0	0	0	0	0
cabaran allah s.w.t terhadap kaum musyrikin mengenai ...	1	0	0	0	0	0	0	0	0	0	0	0	0
ganjaran terhadap orang yg beriman.	0	0	0	1	0	1	0	0	0	0	0	0	0
perumpamaan dan hikmah	0	0	0	0	0	0	0	0	0	0	0	0	0
bukti bukti kekuasaan allah s.w.t.	1	0	0	0	0	0	0	0	0	0	0	0	0
penciptaan manusia dan tugas sebagai khalifah.	0	0	0	0	0	0	0	0	0	0	0	0	1
perintah dan larangan allah s.w.t kepada bani israel.	1	0	0	0	0	0	0	0	0	0	0	0	0
nikmat allah s.w.t kepada bani israel.	1	0	0	0	0	0	0	0	0	0	0	0	0
balasan terhadap perbuatan dan sikap bani israel.	0	0	0	0	0	0	0	0	0	0	0	0	0
pahala bagi orang yang beriman.	0	0	0	1	0	1	0	0	0	0	0	0	0
balasan terhadap bani israel kerana melanggar perjanji...	1	0	0	0	0	0	0	0	0	0	0	0	0
kisah penyembelihan lembu betina.	0	0	0	0	0	0	0	0	0	0	0	0	0
keraguan terhadap keimanan orang yahudi.	0	0	0	1	0	0	0	0	0	0	0	0	0
bani israel mengingkari janjinya dengan allah s.w.t.	1	0	0	0	0	0	0	0	0	0	0	0	0
sikap orang yahudi terhadap para rasul dan kitab yang ...	1	0	0	1	0	0	0	0	0	0	0	0	0

Figure 6: Comparison between Unigram and Al-Humaira
(Sign 1 if got match and 0 if none)

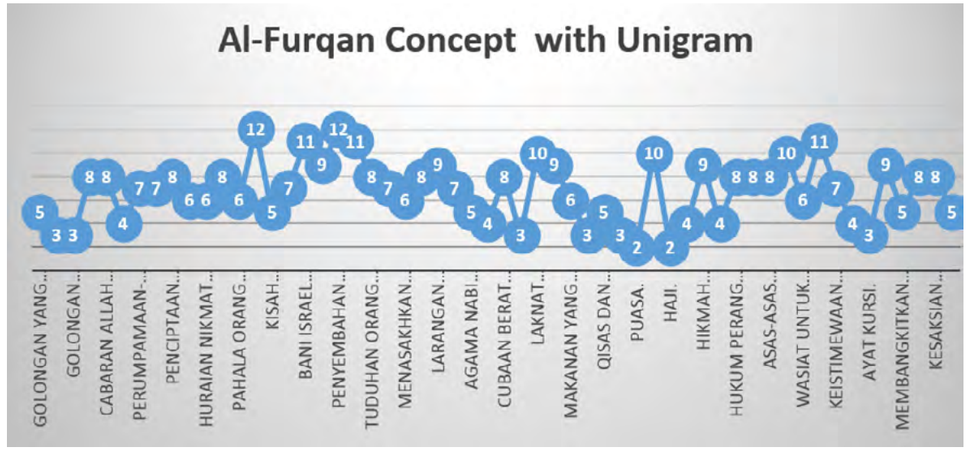

Figure 7: 57 Concept matching in Al-Furqan

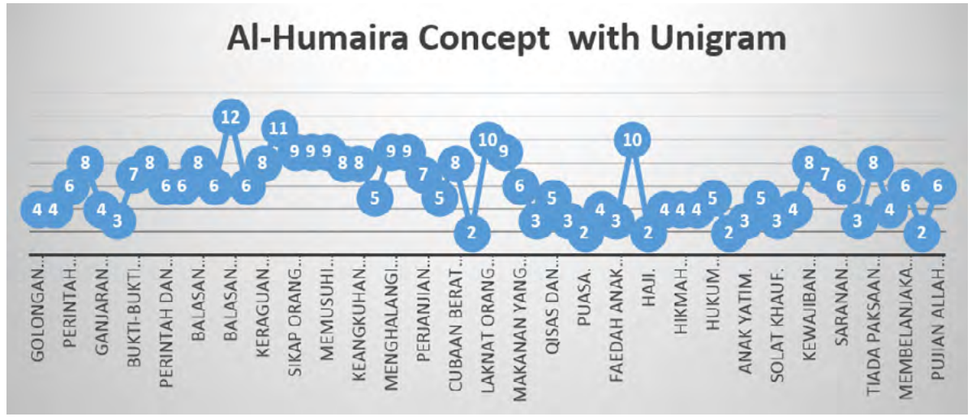

Figure 8: 56 Concept matching in Al-Humaira

Based on the benchmarks, we 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:

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

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:

The benchmarking is prepared with the following datasets:
1. Total of unigram match in Al-Furqan (57 Concept) = 336
2. Total of unigram match in Al-Humaira (56 Concept) = 378
3. Total of unigram result = 1,467 words

Percentage for Al- Furqan
336 / 1,467 X 100 = **22.90%**

Calculation percentage for Al- Humaira
378 / 1,467 X 100 = **25.77%**

The results of benchmark 1 and benchmark 2 show that the matching concept between unigram and Al-Furqan and Al-Humaira was low 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 term matching can be maintained. The similarity of the semantic concept was not covered in this paper.

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 NPL processing such as tokenization, cleaning the text, remove stop words and performed benchmarking for evaluating the similarity of concept with the existing 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 used for learning the Al-Quran knowledge. Future works will explore the using of semantic approach to find a more meaningful concept in 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. https://doi.org/10.14569/IJACSA.2015.060237

Atwell, E., Dukes, K., Sharaf, a., Habash, N., Louw, B., Shawar, B. a., … El-Haj, M. (2010). Understanding the Quran: A New Grand Challenge for Computer Science and Artificial Intelligence. *Grand Challenges for Computing Research. British Computer Society Workshop*, *1*(11), 1829–1841.

A. Fatimah (1995). A Malay Language Document Retrieval System: An Experimental Approach and Analysis in Information Science. PHD. Bangi: Universiti Kebangsaan Malaysia, pp 322.

Atwell, Eric & Dukes, K & Sharaf, A & Habash, Nizar & Louw, B & Shawar, Bayan & McEnery, T & Zaghouani, Wajdi & EL-Haj, Mahmoud (2010). Understanding the Quran: A New Grand Challenge for Computer Science and Artificial Intelligence.

Biru, T. (2015). *Al-Quran Al-Karim dan Terjemahan Al-FAUZAN*. 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

M. N. Kassim, M. A. Maarof, A. Zainal and A. A. Wahab, "Word stemming challenges in Malay texts: A literature review," *2016 4th International Conference on Information and Communication Technology (ICoICT)*, Bandung, 2016,pp.1-6.doi: 10.1109/ICoICT.2016.7571887

Ta'a, A, Abdullah, M. S., Mat Ali, A. B., Ahmad, M. (2014). Themes-based classification for Al-Quran knowledge ontology. International Conference on ICT Convergence. 89-94. 10.1109/ICTC.2014.6983090.

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. https://doi.org/10.7763/IJFCC.2013.V2.229