



Information Visualization for Learning words in the Qur'an

Raja-Jamilah Raja-Yusof^{1,a}, Zulkifli Mohd-Yusoff^{1,b}, Roziati Zainuddin^{2,c},
Mohd-Sapiyan Baba^{2,d}.

¹Centre of Qur'anic Research, University Malaya, 59200, Kuala Lumpur, Malaysia

²Faculty of Computer Science and Information Technology, University Malaya, 50603, Kuala Lumpur, Malaysia.

rjry@um.du.my^a, zulkifliy@um.edu.my^b, roziati@um.edu.my^c, pian@um.du.my^d

ABSTRACT

Qur'an is a source of guidance to Muslims around the world. Although the language of the Qur'an is Arabic, many Muslims with different native languages attempt to learn the language to understand the message of the Qur'an. Qur'an contains many repeated words. Even though there are approximately 77 430 words in the Qur'an, there are only 5155 words used repeatedly or at least once to make up those words. This means that in theory for non-native Arabic speakers, these words can be added to their vocabulary to understand the literal meaning of the message of the Qur'an. In this paper, an interface was developed incorporating parallel and scatter plot visualization to visualize word information in the Qur'an. Users are able to know the Arabic words, the meaning of each word, the word count, word position and their personal vocabulary count to support their learning activity. Interviewed participants found the software useful for learning Arabic words. Although analysis of the two plots show that parallel plot scores better in terms of efficiency of interaction, interviewed participants give lower rating for graphical perceptibility to the plot. The scatter plot visualization is then developed to replace the parallel plot.

Keywords: Qur'an, Visualization, Parallel Plot, Scatter Plot, Learning, Arabic words.

1. INTRODUCTION

Qur'an is a source of guidance to Muslims around the world. Many non-native Arabic speakers have the ability to recite the Qur'an fluently even though they do not understand the language. One of the reasons on how fluency can be achieved during reading or recitation is the fact that Qur'an contains many repeated words. There are 77,430 words in the Qur'an found from the analysis of information gathered through Qur'an.corpus.com by Dukes (2009). Only 5155 of words (or 6.7 % of all words) are used at least once or more to make up all the words in the Qur'an.

Therefore, in theory if non-native Arabic speakers can add these 5155 words to their personal vocabulary, they would be able to understand the literal meaning of the Qur'an. Secondly, for any information to go through to the long term memory (LTM) for learning purposes, a stimulus must attract the attention of one of the sensory system. Therefore, a type of stimulus could be introduced to Qur'an reciters to achieve literal understanding of the Qur'an. This paper proposes visualization as means of attracting the sensory system as well as means of interaction between users and words in the Qur'an for non-native Arabic speakers to learn the literal meaning of Qur'anic words. The strategy is to make known to users the word count, the meaning, the position of each Arabic word and users can track their personal vocabulary list within the interface system.

For example, users might be more interested in learning words that occur more frequently. When users see the visualized information of word frequency in the Qur'anic verses, they might wonder what those words are and where those words can be found. At this point users start to search for the word and try to learn the meaning of the word. Subsequently, a learning process occurs and if this process is repeated the information can go through the LTM. There are many techniques used in information visualization in various application domains that can be categorized into quantity and connectivity (Yusof, Zainudin, Baba, & Yusoff, 2009). Therefore, the visualization technique chosen to attract attention of the sensory memory must support both quantity and connectivity so that at least 4 variables can be displayed (Yusof, Zainudin, Baba, & Yusoff, 2011; Yusof, Zainudin, Baba, Yusoff, & Jomhari, 2012). In our case we have chosen the parallel and the scatter plot implemented only using the 30th Juz (*Juz Amma*) and Sura *Fatiha*. It is implemented for the Malay Muslim community. The novelty of this system is the visualization using parallel and the scatter plot for learning Qura'nic verses for the Malay community as no such system is available in the market.

The next sections of this paper are as follows; the related work, the analysis, design and implementation section, the user study via interview section and finally the discussion and conclusion section.

2. RELATED WORK

Abdulazeez (2004), Abdelbaki (1983) and Yusuf (1992) had attempted on the teaching to understand *Qur'an* based on the most frequently occurring words. Abdulazeez (2004) and his team developed a software for this purpose although he started with designing lesson materials that were posted through email for interested individuals. The software can be downloaded from the UnderstandQuran.com (UQ) website¹. This game software provides 15 *Qur'anic* word lists to be learned and memorized by the user with around 120 frequently used words in the Qur'an. It gives a word by word meaning to the Arabic words listed in the software. This includes breaking the words into segments when a token (a phrase that appears as a word) contains more than a word. Users need to answer questions based on the words provided. The software keeps track of the scores attained to monitor performance. Another such lesson can be obtained from Quraanic Lesson website (QL)². The lessons provided in this website are similar to UnderstandQuran software with additional audio recitation feature provided. Both of these software are for learning Qur'anic words based on Qur'anic words frequencies. One major limitation of this software is that the lessons designed are limited to questions set by the developer of the software. Once finished, users can only repeat and new questions can only be included in the new version of the software. There is also no clear evidence for the effectiveness of the software since no reported user study can be found.

Another website related to understanding Qur'an is the Qur'an Corpus (QC)³. It is a more comprehensive website about the Qur'an and also covers information on Qur'anic grammar. It also include free downloadable Qur'an ontology scripts, useful Java codes and links related to downloadable Qur'anic Arabic and translation databases in many different languages (Dukes , 2009a). Although word frequency is provided in this website, it is not designed for learning purposes; it is more towards providing information to users. A limited visualization view of the Qur'an ontology is also provided though its usefulness in context of understanding the Qur'an is not clear. Another type of Qur'anic software is the Zekr Qur'an (ZQ) software (Saboorian, 2010) and the Tanzil.info (TI) website that provides a computer

¹ <http://www/understandquran.com>

² <http://quraaniclessons.com/index.asp>

³ <http://corpus.quran.com>

interface for reading or reciting the Qur'an using a computer. This type of software is very common and many others develop similar ones providing the facility to search words in the Qur'an, navigate through the book *sura* by *sura*, *ayah* by *ayah*, provides audio recitation from various famous reciters and translations in various languages. However, the software do not provide the support for learning the Arabic language other than providing the translation. Visualization is usually not provided to aid in the learning process.

There are also other English medium software developed mainly for the purpose of improving one's vocabulary. One such software is the Ultimate Vocabulary (UL) software (eReflectSoftware, 2009). It provides usage examples, word visualization, translation in other languages, test, and games and it prepares interested individuals for exams such as GRE and TOEFL. The Free Reading (FR)⁴ website is a website developed specifically for teaching children to read in English (Toppo, 2007). Although it is not a software system, the website is developed in a systematic manner providing activities, instructions, guidelines and resources on how to enhance children's reading skills. Word Recognition is part of its activity apart from offering Letter Sound, Letter Writing, Phonological Awareness and Sounding Out activities. Readily made available are the resources that can be downloaded and used with specific instruction on how to conduct a game / teaching activity.

Another useful analysis is on Arabic language learning software. One of the most comprehensive is the Mualimi⁵ software (M). There many good features of this software and a few of the most outstanding features are such as containing over 75 000 Arabic words with very rich audio-visual presentation, allowing users to plan their study at their own pace and level. Games and play words are also included in the software. Another example is an online Arabic software called Madinah Arabic online tuition⁶ portal (MA). This portal provides an e-book for a comprehension lesson of the Arabic language dividing the book into two categories Arabic reading and Arabic language. The language lesson can be considered complete covering all necessary topics including morphology and verb structure.

3. ANALYSIS, DESIGN AND IMPLEMENTATION

3.1 Analysis

Table 1 shows the summary of features in all the software mentioned in the previous section. There are eleven main features that we extract for discussion; the morphology lesson, visualization, personal achievement, word meaning, educational purpose, games, frequently used words, audio, translation, root word search and Qur'anic lessons. Three Qur'anic based software provide word by word meaning for educational purposes while only QC provide visualization. For non Qur'anic software, the UL provides the most features including visualization and word by word meaning. Although personal achievement is important to be considered as a feature only UQ, QL, UL and M provide this feature. Although the MA provides very comprehensive course structure, it does not use the approach of visualization and frequently used word. Base on this analysis, functional module such as discussed in the next section is proposed with 2 features i.e. the games and audio not included in the current development.

⁴ http://www.freereading.net/index.php?title=FreeReading_Media

⁵ <http://www.arabicteacher.com/arabicL3software/reviewerstnotes.htm>

⁶ http://www.madinaharabic.com/Arabic_Language_Course/Lessons/

Table 1: Comparison of features in Qur’anic software

	Feature \ Systems	UQ	QL	QC	ZQ	TI	UL	FR	M	MA
1	Morphology	√	√	√						√
2	Visualization			√			√			
3	Personal achievement	√	√				√		√	
4	Word by Word Meaning	√	√	√			√	√	√	√
5	Educational purposes	√	√	√			√	√	√	√
6	Games	√					√	√	√	
7	Frequently Used Words	√	√	√			√			
8	Audio		√		√	√	√	√	√	√
9	Translation in other language	√	√	√	√	√	√		√	
10	Root word search					√	√			
11	Qur’anic lesson	√	√	√	√	√				√

3.2 Design and implementation

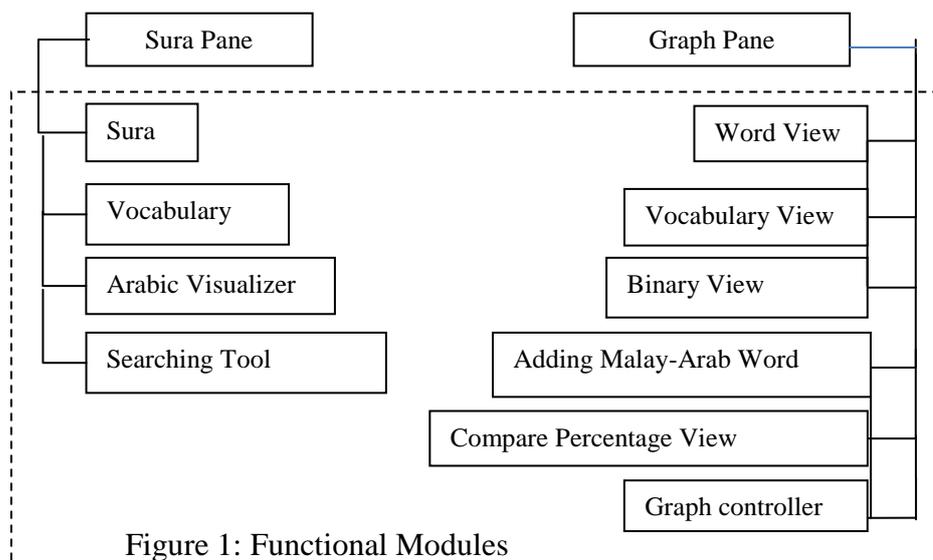
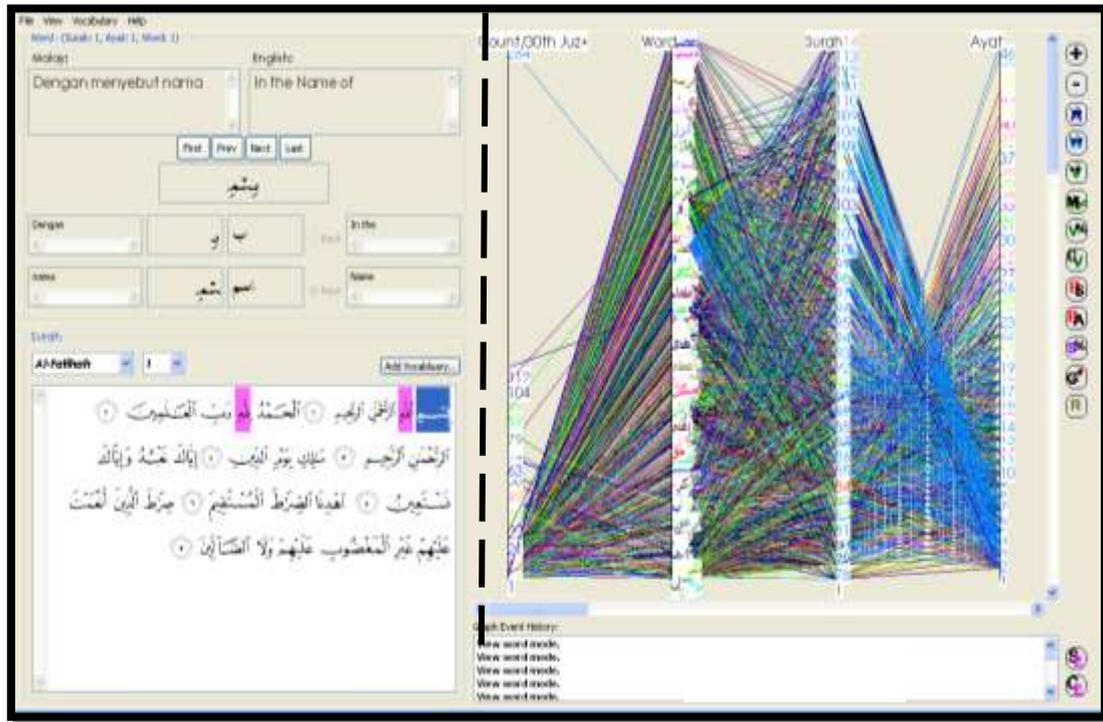


Figure 1: Functional Modules

The interface was implemented on a Dell, Vostro 1200 (T5670) notebook with Intel Core 2 Duo CPU and 1GB Memory. The software environment is based on Windows XP SP3 with Eclipse version 3.3.2 as the Integrated Development Environment. The Java language is used as the implementation language based on Java SE Runtime Environment version 1.6.0_02. MySQL is used as the database. Figure 1 shows the functional modules of the visualization software developed. There are 2 main modules, the Sura Pane and Graph Pane. The Sura Pane contains the Sura, the Vocabulary, the Arabic Visualizer and searching tools while the Graph Pane contains the Word View, the Vocabulary view, the Binary View, Adding Malay-Arab Word View, Compare Percentage View and the Graph controller module.

Figure 2 shows the snapshot of the Qur’anic visualization software using parallel plot. Users are able to learn the meaning of words through the Sura Pane. Here, users are also able to explore word by word meaning of Sura Fatiha and *Juz Amma*. At the same time, they can add to their own vocabulary list, words that they considered “*known*”. In the Graph Pane users are able to detect the word information through the visualization of the word

information (word, word count, word position). They are also able to view their own vocabulary word information mention previously and in terms of percentage in each sura.



4. USER STUDY: INTERVIEWING THE PARTICIPANTS

This study adopts a qualitative approach to find out user's opinion of the interface. The objective of this interview was to find out the perception of users of the interface developed and to get their opinion on the usefulness of the interface for learning the literal meaning of words in the Qur'an. There were four participants involved. 3 of them were aged between 20-28 years and one of them was 40 years old. There were 2 females and 2 males. 3 of them considered themselves as intermediate level Arabic speakers and one of them considered himself as advanced. Two of them possess intermediate computing skill level and two of them were advanced. Their levels of education were Diploma, Bachelor and Master (2 of them). The procedure of the study started with brief introduction to the interface and the purpose of the software is to help non-native Arabic speakers to learn Arabic words in the Qur'an. They were asked a few questions with three main issues:

- i. Learning: Do you think that the software is useful in helping to learn words in Qur'anic verses?
- ii. Parallel Plot: What do you think of the parallel plot before and after explanation of its usage? Do you have any suggestions regarding this issue?
- iii. Function: Can you read the *sura* number on the bottom of the *graph*? Do you think that the zooming function and the text tooltip are helpful in reading the cluttered information?

Table 2: Participant's respond to question

Participant 1	
Q1	- PaCQ is very useful for the purpose of word recognition and if use many times can help in comprehension. -It is also helpful for researchers of Qur'an
Q2	- parallel plot give cluttered view, however if know its usage it is not hard to use - suggest to give viewing information in list form - visualization is useful in this software but suggest other forms
Q3	- sura number from 80-89 can be read correctly - font size use is sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information view
Participant 2	
Q1	- PaCQ is useful for the purpose of word recognition and if use many times can help in comprehension. - it helps us to focus on unknown words since we can identify them easily
Q2	Parallel plot give cluttered view however if know its usage it is not hard to use - visualization is useful in this software but suggest other forms
Q3	sura number from 80-89 can be read correctly - font size use is sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information
Participant 3	
Q1	- PaCQ is useful for the purpose of word recognition and if use many times can help in comprehension. - useful for people who tries to memorize the Qur'an - word count is useful information to help strategies learning
Q2	- parallel plot give cluttered view, however if know its usage it is not hard to use - show word count and word graph only, position in another graph
Q3	sura number from 80-89 can be read correctly - font size use is sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information
Participant 4	
Q1	- PaCQ is useful for the purpose of word recognition and if use many times can help in comprehension. - useful for people who tries to memorize the Qur'an - word count is useful information to help strategies learning
Q2	- parallel plot give cluttered view, however if know its usage it is not hard to use - show word count and word graph only, position in another graph
Q3	sura number from 80-89 can be read correctly - font size use is sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information

5. RESULT AND DISCUSSION

Table 2 shows the summary of participant's response to the questions asked during the interview. Response to question 1 shows that all four participants agree that the visualization interface is very useful for the purpose of learning words in the Qur'an and if use many times can help in literal understanding of the Qur'an. Participant 2 points out that the interface can help to focus on unknown words since we can easily identify them. Participant 3 and 4 say that the software is useful for memorization and Participant 1 says that it is useful for Qur'an researchers. Participant 3 and 4 also agree that the word count information is useful to be displayed to strategies word learning. Response to question 2 shows that all participants perceive the parallel plot as very cluttered (not graphically perceptible plot), however if its usage is known it is not hard to use. All of them suggested other viewing techniques. Participant 1 suggested showing the information in a list form. Participant 3 and 4 suggested showing word count and word graph only, the position was suggested to be displayed in another graph. Response to question 3, shows that all participants read the sura number

correctly. The zooming, binary and tooltip function was found to be useful to make clear the cluttered information.

In response to question 2, an improved interface was developed for the purpose of learning the Qur’anic words. Figure 3 shows the new interface. The parallel plot has been replaced by a similar plot to the scatter plot in the Graph Pane section. Here the users are shown each word within a particular sura. Each time a word in the Sura Pane is clicked, all of those words in that particular sura and in all other sura are highlighted. At the same time the word count is displayed in the next column. In the visualization interface users are able to learn words in the Qur’anic verse. The scatter plot visualization of the word count, word position and percentage information are to support the learning process. Users for example can start to explore the more frequent words which can be found easily via the visualization plot. The information of the word frequency can easily attract the visual memory compared to list of statement written in text form. They are able to strategies their learning by starting to learn the frequently occurring words shown through the visualization plot. New words can be added to their personal vocabulary list.

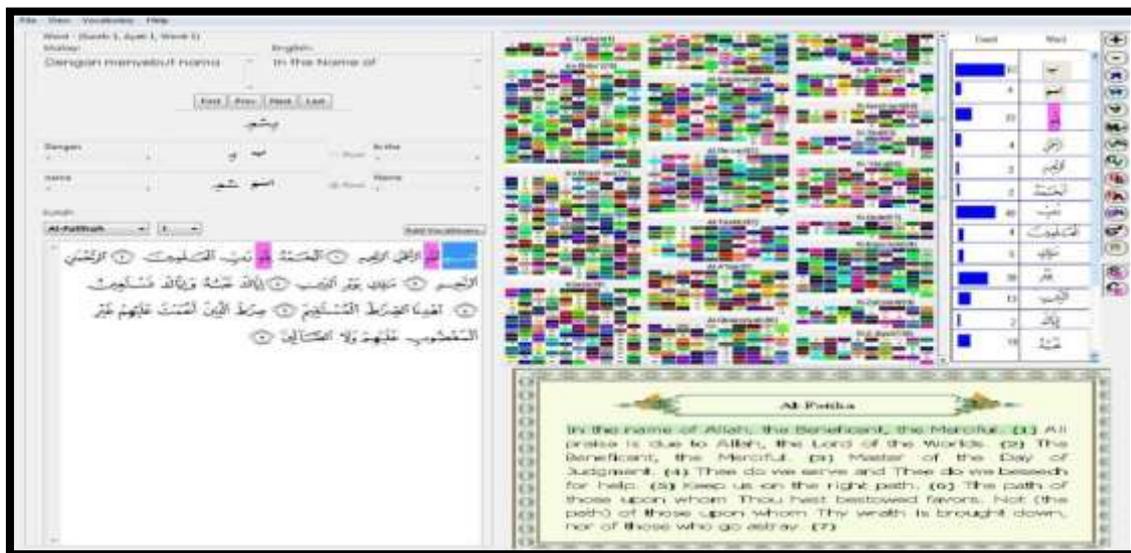


Figure 3: Scatter plot visualization of Qur’anic

Alternatively, users can also learn the words as they read a *sura* or *ayah* through the semantic word visualizer and subsequently add those words involved in the personal vocabulary list. The growing vocabulary list can be used to further strategies the learning by analysing, interpreting and deducing the vocabulary percentage in relation to all *suras*. The next step is to focus on unknown words. Users can actively decide which *sura* to read and learn so that many words in the Qur’anic verse can be recognized. Other functions such as comparing percentage, searching, and filtering, also helps in the rehearsal process of learning and therefore provide active learning and if repeated frequently should make users able to achieve literal understanding of Qur’anic verses.

6. CONCLUSION

This study adopts visualization using the parallel and scatter plot as means to support the learning process of Qur’anic words to achieve literal understanding those words. The

learning strategy suggested by the software is to start with the more frequently repeated words since Qur'an contains many repeated words. The user study via interview found that users agree that the interface is useful for learning the Qur'anic words. However, users are not inclined to the parallel plot visualization as it perceived as cluttered and confusing. Therefore, the scatter plot was used to replace the parallel plot.

7. REFERENCES

- Abdelbaki, I. (1983). Teaching Arabic through the Holy Quran to other than Arabic speakers, the medium level. Master Thesis for the International Institute of Khartoum for Arabic Language for non-Arabic speakers of The Arab League Educational, Cultural and Scientific Organization.
- Abdulazeez, A. (2004). *Understand Qur'an the easy way*. Hyderabad: Understand Qur'an Academy.
- Dukes(a), K. (2009). *Phonetic Transcription*. Retrieved June 2010, from Qur'an Corpus: <http://corpus.quran.com/documentation/phonetic.jsp>
- eReflectSoftware. (2009). *Ultimate Vocabulary*. Retrieved August 3, 2010, from <http://www.ultimatevocabulary.com/>
- Saboorian, M. (2010). *Zekr - Quran Study Software for Windows*. Retrieved 2012, from The Zekr Project: <http://zekr.org/quran/en/quran-for-windows>
- Toppo, G. (2007). *Free online materials could save schools billions*. Retrieved 2012, from USA Today: http://www.usatoday.com/news/education/2007-11-06-freereading_N.htm
- Yusof, R., Zainudin, R., Baba, M., & Yusoff, Z. (2009). Visualization Systems Supporting the Reading of Arabic Document for non-Arabic speakers. *Information Technology Journal*, 16-27.
- Yusof, R., Zainudin, R., Baba, M., & Yusoff, Z. (2011). Parallel versus Perpendicular Plots: A Comparative Study. In A. A. al. (Ed.), *ICIEIS 2011, Part IV, Communications in Computer and Information Science (CCIS)*, (pp. 436-448). Kuala Lumpur.
- Yusof, R., Zainudin, R., Baba, M., Yusoff, Z., & Jomhari, N. (2012). Visualizing Qur'anic words with parallel plot. *Malaysian Journal of Computer Science*, 90-106.
- Yusuf, L. M. (1992). The design of a method of teaching Arabic of related purposes (the understanding of Quran and Hadith), for primary level in Malaysia. *Master Thesis for The International Institute of Khartoum for Arabic Language for non-Arabic speakers of The Arab League Educational, Cultural and Scientific Organization*.