

Natural Language Processing Approach for Searching the Quran: Quick and Intuitive

Abidah Zainal, Noor Azah Samsudin¹, Mohammad Airin Azman, Aida Mustapha

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat, Johor, Malaysia

Abstract. The Quran is a scripture that acts as the main reference to people which their religion is Islam. It covers information from politics to science, with vast amount of information that requires effort to uncover the knowledge behind it. Today, the emergence of smartphones has led to the development of a wide-range application for enhancing knowledge-seeking activities. This project proposes a mobile application that is taking a natural language approach to searching topics in the Quran based on keyword searching. The benefit of the application is two-fold; it is intuitive and it saves time.

1 Introduction

The Quran is a book of guidance for all people all over the world. The Quran covers all aspects from economy to political, science, to astronomy. It has vast information from the beginning of time to the end of time. The Quran consists of 30 Juz, which was descended to Prophet Muhammad S.A.W in 23 years. The Quran serves as a record of time and act as it is most important source in tracing the development of Islam from Mecca to Madinah. Until today the Quran holds as the most trustable book in the world, therefore, the Muslim community regards the Quran as the highest literature for Arabic language. Many miracles in the Quran can be found such as the development of human embryo which recorded in Quran around 1400 years ago, which modern civilization found it in early 1970. Quran has also become as a guideline to time in history, which records actions that are precise and proven as a helping tool for anthropologist [1].

At the same time, Quran has proven to have no errors despite its diverse topics with great depth of knowledge. Based on this characteristic, a searching application for the Quran is proposed. The searching ability is based on natural language input and will be useful to shorten the time taken to search in the entire Quran. Although the digital Quran is convenient to perform searching because a normal search function is sufficient, the source of search is limited. This is because the knowledge is not comprehensive and only covers the searched keywords. The objective of the proposed application is three-fold; (1) to design a Quran search application based on keywords entered in natural language, (2) to develop the application based on mobile platform, and finally (3) to test the application

¹ Corresponding author: azah@uthm.edu.my

usability among the religious officer at Islamic Centre at Universiti Tun Hussein Onn Malaysia (UTHM).

The application uses tokenization to extract keyword when the user enters a sentence. The knowledge returned to users is based on the keywords and covers the tafsir, hadith and translation from a specific verse. The knowledge base is limited to the 30th surah in the Quran known as Juz Amma. The target users are common users with no age restriction. The application, however, will be evaluated by the Islamic knowledge expert at UTHM. This application is hoped to change the way users interact with the content of the Quran and attract young users to learn Quran in a more convenient manner. This application also provides the user specific information about Quran, complementary forms which are essential in finding information in urgent matter

The rest of the paper is organized as follows: Section II describes the related work on mobile-based Quranic applications, Section III presents methodology, Section IV presents the prototype implementation and testing, and finally, Section V concludes the work and highlights future research.

2 Related Work

This section reviews applications related to the Quran discovery using natural language processing. The review is important in order to understand how the system works and how it can be improved for future purposes.

2.1 iQuran Application by Isysway [2]

This application is invented to help user to read the Quran. It is very interactive which has bookmark for last recitation of the Quran. This Quran application by Isysway is available to download at Playstore for free. It also provides searching mechanism to search for verses in each surah that the users want to read. The searching function is using hardcode to display the result of the search and it is similar to the traditional method of reading the Quran but using mobile phone. Figure 1 shows the graphical user interface (GUI) of the iQuran application.

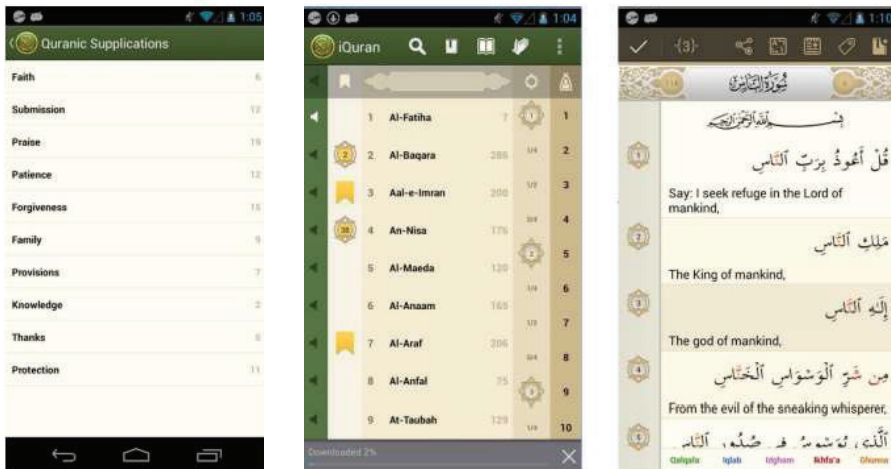


Fig. 1. Graphical user interface of iQuran.

2.2 Quran for Android by Quran.com [3]

Quran for Android is a free source application that available for android users to download using Google PlayStore. The main application functionality is to read the Quran in smartphone. The application requires Internet connection for user to download audio for the application. Due to this limitation, the application have problem in stopping the audio of the Quran when there are a call on the user phone. The application has a night mode that helps to lessen the stress of eyes when reading a text from a bright screen in dark area. It also uses bookmarks to mark the last verse of reading. This is to ensure users reads based on the correct continuity of the Quran. Figure 2 shows the interface for the Quran for Android application.

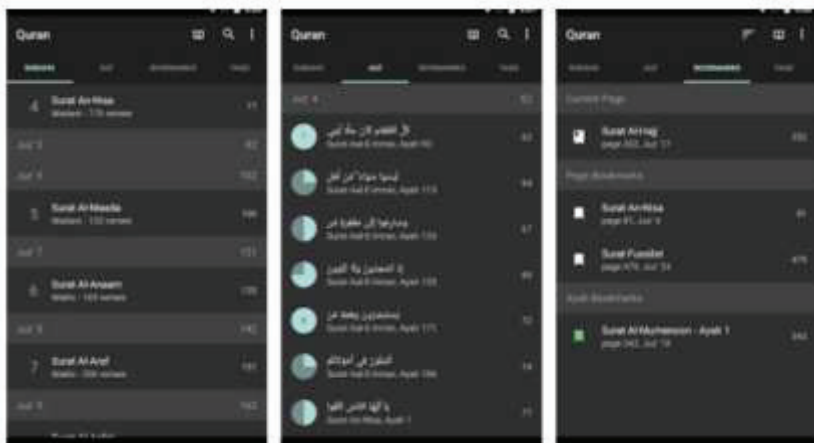


Fig. 2. Graphical user interface of Quran for Android.

2.3 Learn Quran by Badr Interactive [4]

Learn Quran is designed for users who want to start learning the way to read the Quran. Since learning Quran is very hard and some people have limited time on it, the application is developed to countermeasure this problem. This application aids the user to learn the Quran from the pronunciation to tajwid, with audio in helping the user to achieve the correct pronunciation for each verse. Although this application is free, it has premium feature which the user can get when paid for the application, such as no advertisement and more learning tutorial about the Quran. Figure 3 shows the user interface for the Learn Quran application.

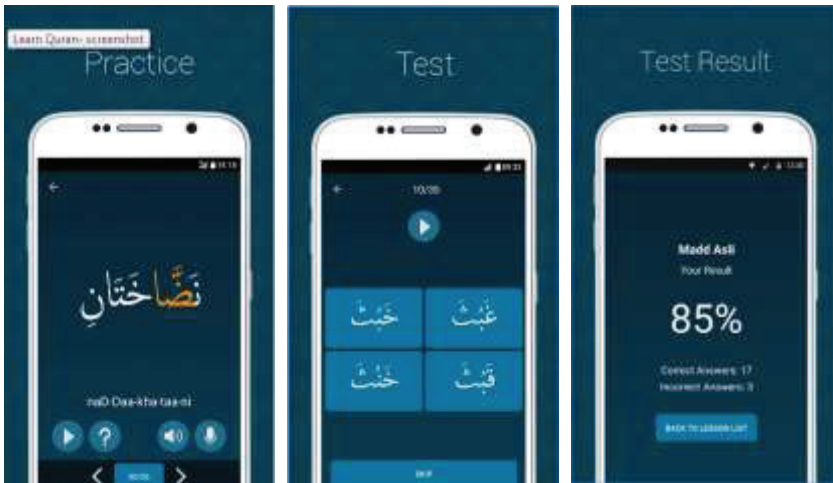


Fig. 3. Graphical user interface of PexPix.

The reviews concluded that all the proposed Quran Search Tool using Natural Language Processing offers a unique feature that can help user to enter input in a full sentence and the tool can still give answers to the search query.

3 Methodology

Methodology is an essential element for developing a work plan for the project in order to establish a mechanism to coordinate and control the flow of the project, and to ensure the project are finish within time. Research has shown that there are many techniques and models available for development of application such as the waterfall model, the prototype model, and the spiral development model. This project uses the prototype model to develop the proposed Quran Search Tool using Natural Language Processing. Figure 4 shows the steps in the prototyping model [5].

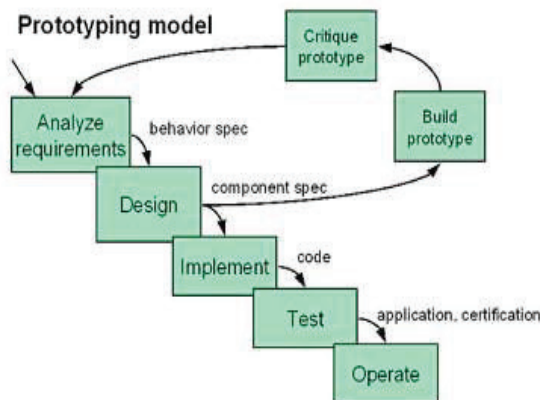


Fig. 4. Prototype development model.

The prototype model was chosen because the technique and flow of this model are repeatable, therefore allowing flexibility and efficiency. This is because the prototype will allow the users to have a hands-on experience of the product before it finishes. This will be a major boost in helping the users to understand what the application is capable of. At the same time, the users can request for changes since they already have a glance on the application before it is launched. The prototype model has six main phases needed to develop the Quran Search Tool using Natural Language Processing, which are planning, analysis, design, prototype development, and testing. The planning phase is a critical stage which will determine the outcome of this project, in this phase, the objective of this project will be determined, as well as model that suitable for the proposed project to be used.

Multiple method and research techniques have been used in ensure the development of this project is on the right track and while at the same time meets the requirements from the used. The activities that will be performed during this phase is to study Final Year Project Thesis from past year students, discuss with supervisor and review of the Internet. Gantt chart will be made to ensure the sustainability of this project running smoothly according to its schedule. At the same time, the project requirements were gathered to make the project a success. Analysis is a crucial which takes a lot of time such as identifying the hardware and software needed to make this application a success.

The goal of this phase is to provide the development of the application with much more understanding of the process to develop a prototype of the Quran Search Tool Using Natural Language Processing. The minimum requirements in this particular phase were collected from the interview with the religious officer in UTHM Islamic Centre. Furthermore the suggestion and recommendation from the center will be taken as guidance for the application. The traditional searching method in existing Quran application requires a longer period of time as compared to the natural language processing techniques that takes less time and makes the application more exciting.

In the analysis phases, the interface of the application will be designed, the navigation will be determined, the button on the application will be designed using Adobe Photoshop to make it more interactive for the users. At the same time, the flow and the way the application works will be specified using the Use Case Diagram, Flow Chart, and Activity Diagram. The design process of this application has been made using multiple hardware and software that are suitable for the development and success of this application. Implementation phase is where the prototype is developed and tested based on design, testing process need to be done before the launch of the proposed Quran Search Tool using Natural Language Processing application. One of the main reason to have a testing procedure is to ensure the application is able to perform according to what has been expected during the planning process, At the same time, in this phase testing is also carried out to remove any bugs or issues according to the scope and the application objective of the project.

4 Prototype Implementation

This section focuses on the analysis and design phases of Quran Search Tool using the Natural Language Processing. A research has been made to ensure the outcome of analysis and designed activities are based on the requirement of this project. At the same time discussion also act as important role to the overall project before the development. Requirement specifications of application need to be analyzed in order to make sure the details are correct, the output to meets the goal and objectives. Before the development process, the application flow needs to be designed based on information from the interview. The analyses also focused on the chosen keyword from the users, the most favorable

keyword in Quran, most common keyword are also identified, so that the project will be able to assist user in understanding the Quran.

4.1 Natural Language Processing

Natural Language Processing is an algorithm that manipulates string to maximize its use in retrieving data. Natural language processing has two primary roles to play in the storage and retrieval of large bodies of information: providing a friendly, easily-learned interface to information retrieval systems, and automatically structuring texts so that their information can be more easily processed and retrieved. Figure 5 shows the sample code excerpt from Android Studio IDE that describes Natural Language Processing.

```
input = (EditText) findViewById(R.id.editText);
String str = input.getText().toString();
bilangan = (TextView) findViewById(R.id.textView6);
no_surah = (TextView) findViewById(R.id.textView2);
kontent = (TextView) findViewById(R.id.textView4);
hadis = (TextView) findViewById(R.id.textView3);
String result = str.replaceAll("[\\-\\+*\\.%\\?| ]", "");

textout = (TextView) findViewById(R.id.textView2);
String trim= str.replaceAll("[\\-\\+*\\.%\\?|>|<| ]", "");
String[] s = trim.split(" ");
```

Fig. 5. Natural language processing.

Natural Language has been involved in numerous systems and games. One of Natural Language Processing's feature is removing unnecessary word, special character, and even symbols. This makes user interaction much more interactive since when user enters symbols the application still recognizes it requires a keyword, second is word separation the application can receive many inputs from user such as a sentence this way the application is capable of chatting with the user. Natural Language Processing also involves artificial intelligence for retrieving data according to specific keywords. Quran Search Tool Using Natural Language Processing uses trim function from Natural Language Processing, word separation and matches keywords to display the correct answer according to user input.

4.2 Use Case Diagram

Use case diagrams are useful in explaining the functionality of the system itself. In the development process, use case diagrams act as a reference for the activity in the system. The use case diagram contains an actor, relationship, and use cases. Figure 6 shows the Use Case Diagram for the Quran Tool using Natural Language Processing.

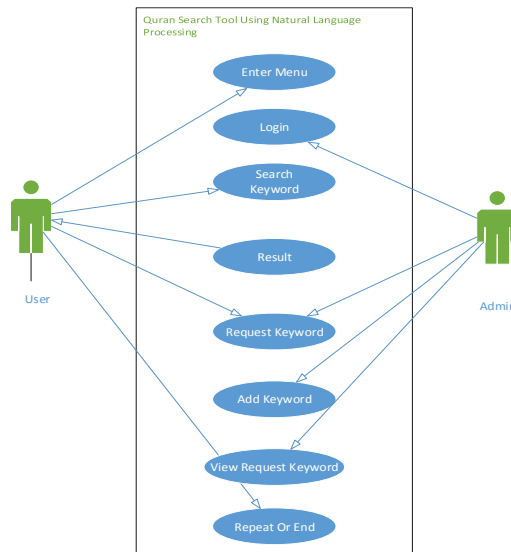


Fig. 6. Use case diagram.

4.3 Sequence Diagram

Sequence diagram for the proposed application shows the steps when the user enters the input sentence, which has a correct keyword. The system will return a result from the application and the user can continue session by entering new sentence or the user can end session. Figure 7 shows the sequence diagram for the entire system.

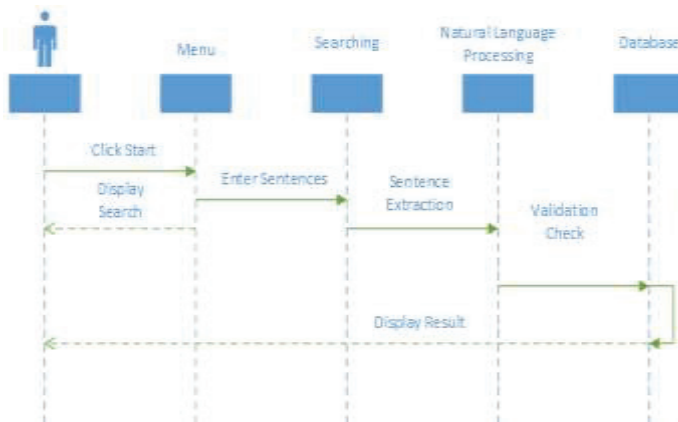


Fig. 7. Sequence diagram.

The application will send back a message shown that they keyword inserted exist. Meanwhile, the user can enter a new sentence to continue search or end the session.

4.4 Activity Diagram

Activity diagram acts as a guide to understand the flow and process of the application. The activity starts when the actor, which is the user, begins the application. After that, the user will enter keyword that they want and the natural language processing module will start. The natural language processing module is important to extract the correct keyword from the database. If the user enters a wrong keyword, no answers will be displayed. The user has a choice whether to start again the search or end the session. If succeed, the application will display the result of the keyword entered in another interface. The result will be generated on the page.

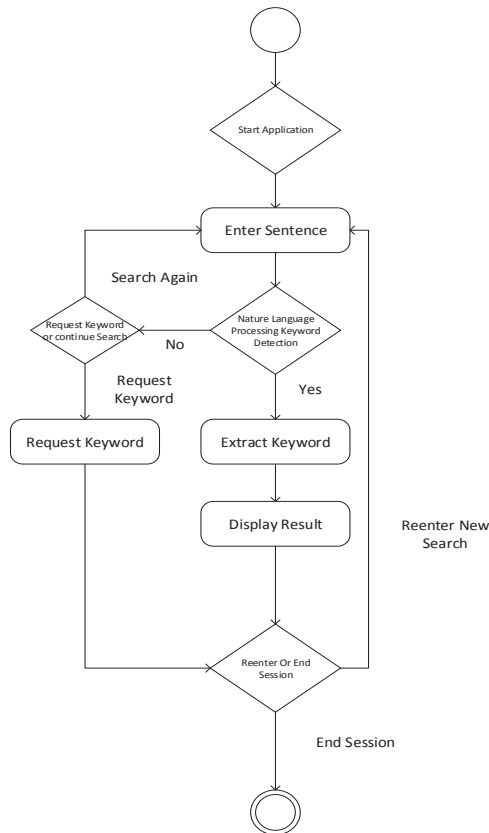


Fig. 8. Activity diagram.

Figure 9(a) illustrates the start interface of the application. The start interface consists of one button which consist a function to enter another activity which is the searching activity. Once the user click starts, the search interface will appear as shown in Figure 9(b). The search menu of this application consists of two buttons (1) input from user which is a sentence, and (2) text that will view the result of the searches. The button which has the word “CARIAN BARU” acts as a clear button if the user wants to clear the search sentence input by user. The green button function is to grab the sentence and initialize a process to generate the output for the two texts which will view.

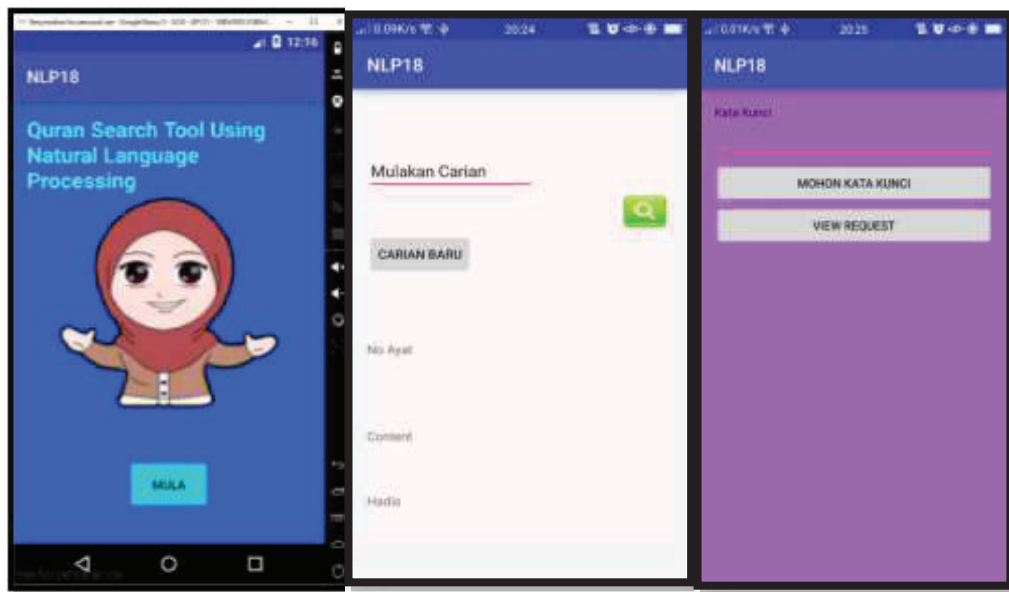


Fig. 9(a). Welcome page.

Fig. 9(b). Search page.

Fig. 9(c). Keyword request.

5 Conclusions and Future Works

The project started with the object-oriented analysis which defined all the requirements of the system. Then the object-oriented design phase which the application displays the main menu of the application, search interface, popup when a keyword input does not match, and a database which integrated from android to the website. Then enters the implementation and testing phase where the system is built and tested according to system requirements and objective, the overall of this chapter will focuses on limitation and future improvement of this application. One important limitation that need to be considered is that the nature language processing module in this application only can match with one keyword. If a keyword consists of space the application will not detected since the application remove the space character.

In order to improve its performance and functionalities, it is recommended to develop this application to other mobile platform such as Symbian for Nokia smartphone, IOS from Apple Inc., Windows phone for HTC and also Nokia from Microsoft, and Blackberry 11 from Blackberry. This will increase the flexibility and amount of users using this application since android only accommodate around 45 percent of all smartphone in this world. Moreover, the usability of this application can be improved by adding other functional ability such as to detect two keywords at the same time. Overall, it is hoped that this application is able to promote the culture of learning the Quran among the Muslim youngsters today by riding on the smartphone technologies.

Acknowledgements

This work was supported in part by a grant from the Ministry of Education under the Research Acculturation Grant Scheme (R045) and in part by a grant from Research Gates IT Solution Sdn. Bhd.

References

1. E. John, L. Esposito, The Oxford History of Islam (1999)
2. iQuran by Isyway [Online] Accessed on 21 October 2015 at <https://play.google.com/store/apps/details?id=com.isysway.free.alquran&hl=en>
3. Quran.com [Online] Accessed on 21 October 2015 at <https://play.google.com/store/apps/details?id=com.quran.labs.androidquran&hl=en>
4. Learn Quran [Online] Accessed on 21 October 2015 at <https://play.google.com/store/apps/details?id=com.bi.learnquran>
5. K. E. Lantz, The prototyping methodology. Prentice-Hall, Inc. (1986)