

Development of Qur'an Search Engine For The Indonesian Language Query

1st Fenty E.M.A¹, 2ndRina Candra N.S¹ 3rd A. Juaini Syukri²

{fenty.eka@uinjkt.ac.id¹, r_candra_ns@edu.unisbank.ac.id¹, juaini.syukri@uinjkt.ac.id²}

Informatics Department, Faculty of Science and Technology, Universitas Islam Negeri Syarif Hidayatullah, Jakarta¹, Faculty of Law and Sharia, Universitas Islam Negeri Syarif Hidayatullah, Jakarta²

Abstract. This study aims to build a search engine for the Qur'an verses of Indonesian keywords that are furnished by thematic index of the Qur'an. The query is inputted using the Indonesian keyword which will be matched with the translation database of Al Qur'an in Indonesian language. The resulting query results shown will provide the highlight on the search keyword, users can save the result to xls and/or pdf format. If users feel the search results are less relevant to the desired theme, they can search based on the list of thematic classifications available in tree form.

Keywords: Qur'an Search Engine, Thematic Verse Classification. Indonesian Query

1 Introduction

The website www.quran.com holds a search feature that can be used to search for Qur'an verses with keyword input in the Indonesian language, but the search results are sometimes less relevant to the purpose users desire [1]. Also search results cannot be directly downloaded. The website www.alquranalhadi.com holds a fairly complete thematic index and allows users to search for the desired Qur'an verse [2]. Based on observations on both websites, a questionnaire was developed to explore the needs of users on a search engine of the Qur'an verses.

The results of the questionnaire distribution showed the results that the main needs of a search engine site of the Qur'an are: (1) users can input the keyword of the verses to search, (2) users can download the search results, (3) users can search the Qur'an verse according to the desired theme, (4) users can search the Arabic root from search results, (5) users get relevant search results, even if they do not contain the keywords entered. Based on the results, the research was conducted with the aim to develop a website that accommodates the main needs.

The research is limited to three main features: (1) searching for Qur'an verses using Indonesian keywords, (2) searching for Qur'an verses based on thematic index, (3) downloading facility of search results. The first feature will be tested for its accuracy and relevance. The second feature will be tested from its usability qualitatively by users. The third feature will be tested from the accuracy between the results of the download with the display on the web.

The Qur'an database and its translation was obtained at <http://tanzil.net/docs/resources>. System was developed using Rapid Application Development (RAD), implemented using PHP and MySQL language. The thematic classification index is built on the Qur'an Syaamil thematic index [4]. The website interface was designed using Task-Centered System Design (TCSD) approach. Downloaded files are stored as xls and pdf files format.

2 Research Development On Qur'an Search Engine

Computational systems for analyzing and explaining the Qur'an on a morphological basis were developed at the University of Haifa. The copy of the Qur'an used is a phonemic transcription of a standard Arabic text. This transcription is based on ASCII notation. This system performs a number of questions on the Qur'an text that makes references not only with words but also for its linguistic attributes. Thus, users can extract from the text of certain words or word patterns, using feature words (such as root, pattern, lexical, gender, number, dependent, tense and aspect, etc.); or a combination of words that fit a particular structure [7]. The Corpus Al Qur'an produced by the University of Haifa is considered a gold standard which is suitable for comparing Arabic morphological analysis results, reformatted and adapted as the gold standard in MorphoChallenge 2009, a research content on morphological analysis.

Another way to explore information in the Qur'an is by stemming the transliteration of the Qur'an verses. The technique used is proven to provide superior results over other stemmers developed for Arabic. By using transliteration the basic Arabic word can be obtained and the frequency of its appearance is calculated. It also facilitates the creation of thematic structures of chapters of the Qur'an. Qur'an verses that have similar themes, are grouped in a cluster and recorded in relation to one verse with another. The construction and interpretation of the cluster tree semantics based on lexical frequencies are proven to be a useful approach for finding thematic links between letters and verses of the Qur'an.

This method is easier to implement than using stemming in its Arabic script. The Qur'an question and answer system of Al Bayan Question Answer accepts inputs in the form of questions about the Qur'an in Arabic, takes the most relevant verses of Qur'an, then extracts the sections containing the Qur'an and Tafseer's answers [8] Mechanisms: (1) Preprocessing Operations; (2) Question Analysis; (3) Information Retrieval (IR); (4) Answer Extraction.

Formal methods are also used in the Qur'an search system. The formal method is a math-based technique for description and verification of software and hardware system specifications. The one of studies on the use of formal methods for Natural Language Processing in the Qur'an Searching System (QSS). Z notation is used to express the formal specification of three QSS search techniques that are text based, stem-based and synonymous based. QSS allows users to search for keywords in the Qur'an and retrieve relevant verses. The Z/EVES tool is used to check and analyze Z specifications. [9] each type of specification is entered and checked one paragraph at a time.

3 System Development Method

On an outline basis, the system development consists of several stages: (1) stage of users need determination, (2) design stage, and (3) implementation stage. The need determination

stage generates three main needs of the search feature of the Qur'an verse. At the design stage an architecture design and system interface are conducted. The design of the system architecture including the design of the Qur'an database. Interface design consists of four stages: (a) identifying tasks, (b) determining users need from tasks, (c) designing interface scenarios, (d) evaluating existing scenarios. Implementation stage is to apply the design with PHP programming language, MySQL database and perform users acceptance test.

4 Results and Discussion

4.1 Requirement Planning

At the stage of determining the needs of users then the results are obtained: (a) Al Qur'an database is required with Indonesian translation from the Indonesian Ministry of Religious Affaris which can be downloaded from www.databasequran.org website. (b) The index and theme classification are selected from Syaamil Qur'an and created one table in the database.

4.2 Workshop Design

The downloaded Qur'an database is modified by adding one table to collect classification index data. The result of the modification is shown in figure 1.

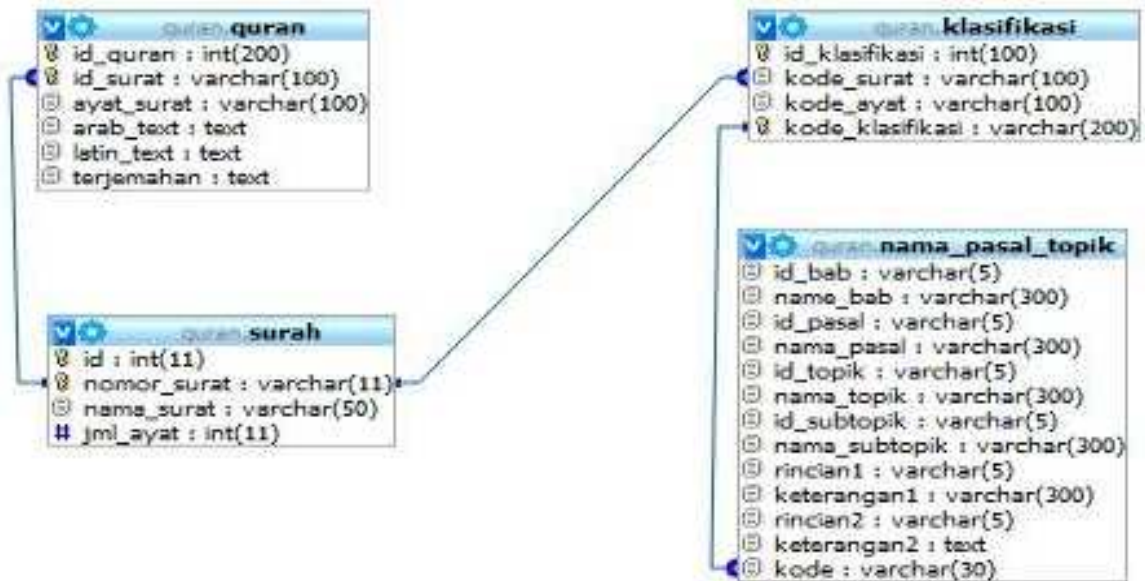


Figure 1. Relational Table on the Qur'an Database

The design result of this system interface is shown in figure 2.



Figure 2. Home Page of Quran Search Engine for Bahasa

KLASIFIKASI AL-QURAN

BERANDA BACA AL-QURAN KLASIFIKASI TENTANG

BAB KLASIFIKASI

bab → [sekitar Arkanul Islam](#)

pasal → [AD DIEN \(AGAMA\)](#)

topik → [Agama yang Diridhai Allah](#)

subtopik → [-](#)

keterangan1 → [-](#)

keterangan 2 → [- Klik Untuk Lihat Ayat](#)

- [Tidak ada paksaan](#)
- [Ajakan Kepada Islam](#)
- [Hakikat Islam](#)
- [Ikhlas dalam Beragama](#)
- [Orang-orang Islam](#)
- [Jahiliyah](#)
- [TALHID](#)
- [MULHAMAD RASULULLAH SAW](#)
- [SHALAT](#)
- [PUASA](#)
- [ZAKAT DAN SHADAQAH](#)
- [HAJI DAN UMRAH](#)
- [BENTUK-BENTUK BADAH](#)
- [IMAN](#)

Tema : sekitar Arkanul Islam → AD DIEN (AGAMA) → → → → →

Al-Bayyinah ayat: 4

وَمَا كَفَرْنَا بِأَلَّهِ إِنَّا مِن مَّجَاهِلِينَ

Wa ma kafarraqi lalldiina uutul kitesba illa min ba'di ma jaraat humul bayyinqh
Dan tidaklah berpecah belah orang-orang yang diatangkan Al Kitab (kepada mereka) melainkan sesudah datang kepada mereka bukti yang nyata.

[008004 Lihat Tafsir](#)

Al-Bayyinah ayat: 5

وَمَا أُمِرُوا إِلَّا لِيُعْبَدُوا اللَّهَ مُخْلِصِينَ لَهُ الدِّينَ حُنَفَاءَ وَيُقِيمُوا الصَّلَاةَ وَيُؤْتُوا الزَّكَاةَ وَهُمْ عَلَىٰ الذِّكْرِ حَافِذِينَ

Wa ma amiruu illaa liya'budulloaha mukhlisihina lahuddiina hunafaa-a wa yuqimush shalaata wa yu'tuzzakataa wa dza'ika d'irul qayyimah
Padahal mereka tidak disuruh kecuali supaya menyembah Allah dengan memurnikan ketaatan kepada-Nya dalam (menjalankan) agama yang lurus, dan supaya mereka mendirikan shalat dan menunaikan zakat; dan yang demikian itulah agama yang lurus.

[0080051 Lihat Tafsir](#)

Figure 3. Thematic Classification

4.3 Usability Implementation and Testing

The design is implemented using PHP programming language with MySQL database. Furthermore, usability testing is performed to ask the representative users to evaluate the performance of the Qur'an search engine. Two participants have been selected, one expert represents the Qur'an (Al Hafidz - memorized 30 juz) and another represents the novice users. Based on the observation, the usability testing results obtained are:

1. The search menu is easy to recognize and keywords are easily inputted.
2. Users easily search with keyword input.
3. The search results are easy to read and downloaded into xls and pdf files format.
4. Search results still have weaknesses where irrelevant verses to the keyword are found.
5. Users easily use the classification menu to search for Qur'an verse themes
6. Users can search for Qur'an verses through the classification menu "Per Juz".

4.4 Recall and Precision Testing

The testing of search engine functional using 8 keywords and focus only the verses in *juz 1 (chapter 1)*. The basis for query retrieval is the Convenience Sampling method, that is, the researcher assumes that the words at the top are often used, while the bottom is rarely used.

Table 1. Recall and Precision Testing

No	Query	Number of Verses	Number of Words	Recall	Precision
1	Israil	50	52	100%	100%
2	Bani Israil	66	53	100%	80,30%
3	Yahudi	37	40	100%	100%
4	Nasrani	22	25	100%	100%
5	Yahudi Nasrani	38	40	100%	100%
6	Baitul	13	14	100%	28,57%
7	Maqdis	3	3	100%	100%
8	Baitul Maqdis	13	3	100%	21,43%

In table 1 it can be seen that the key words of **Bani Israil** produced 66 verses with the correct number of words of the **Bani Israil** as many as 53. There are 2 verses which mention the words of **Bani Israil** and the words of **Israil** together in the same verse. Therefore, when searching with keyword of **Israil**, 50 verses were obtained, with the number of **Israil** words being 52. While the words that contain the word **bani**, but are not relevant to the **Bani Israil** keywords in detail can be seen in the table below:

Table 2. Irrelevant Words Founded by The Keywords of

Bani Israil		
No	Query	Number of Words
1	Membebani	4
2	Rabbani	1
3	Dibebani	2
4	Bani Adam	2
5	Bani Quraizah	1
6	Rahbaniyyah	1
7	Baniih	1
8	Zabaniyah	1
Total Words		13

Likewise, when using **Yahudi** dan **Nasrani** keywords when searching in 1 word, **Yahudi** or **Nasrani**, get recall and precision 100%. However, if you use **Yahudi dan Nasrani** keywords, the system will display 4390 verses. This is due to the word and which is included in the search term. words **dan** is not sought after by word but become a unity in **Yahudi dan Nasrani** phrases.

The keywords **Baitul Maqdis** produces 13 verses, but these verses only contain 3 phrases of **Baitul Maqdis**. This is contrary to the keyword Baitul, which produces 14 words of **Baitul** which are contained in 13 verses. The verses, 10 words of **Baitul** are taken from pieces of the word **Baitullah**. This is not relevant to the purpose of the keyword **Baitul**. Number of verses containing the phrase **Baitul Maqdis**. This causes precision values of only 21.43%.

5 Conclusion

The created search system can already be used by users to search for Qur'an verses based on Indonesian keywords. Search results can also be downloaded to xls and pdf files format. The search results still contain the sections that are less relevant to users intent. This is due to a search that only searches for a match between a keyword and an existing word in the database. All menus have been performed from usability testing, and meet the requirements of users. The system has produced the correct query answer, this is indicated by the recall value of 100%. but the precision is still low, below 30%.

6 Future Work

The created search system can already be used by users to search for Qur'an verses based on Indonesian keywords. Search results can also be downloaded to xls and pdf files format. The search results still contain the sections that are less relevant to users intent. This is due to a search that only searches for a match between a keyword and an existing word in the database. All menus have been performed from usability testing, and meet the requirements of

users. The system has produced the correct query answer, this is indicated by the recall value of 100%. but the precision is still low, below 30%.

Acknowledgements. This research was funded by Center of Research and Publication (Pusat Penelitian dan Penerbitan - Puslitpen) UIN Syarif Hidayatullah Jakarta. Thanks to Rifda Younanda Mutiara for research assistance.

References

- [1] www.quran.com diakses tanggal 05 sampai dengan 12 juni 2017, 09.00 – 17.00.
- [2] www.alquranalhadi.com diakses tanggal 05 sampai dengan 12 juni 2017, 09.00 – 17.00.
- [3] <http://tanzil.net/docs/resources> diakses tanggal 15 Januari 2017, 21.00.
- [4] Syaamil Qur'an, "Terjemah Per Kata", Mizan Group, Bandung, 2010.
- [5] N. H. Abbas, "Quran 'Search for a Concept' Tool and Website," 2009.
- [6] H. Abdelnasser, "Al-Bayan : An Arabic Question Answering System for the Holy Quran," 2014.
- [7] A. Al-Gharaibeh, A. Al-Taani, and I. Alsmadi, "The Usage of Formal Methods in Quran Search System," ICICS 2011, 2011.