Implementation Levenshtein Distance Algorithm for Hifdzil Quran Quiz

Wahyudin Darmalaksana[1], Ali Rahman[2], Egi Andriana[2], Ichsan Taufik[2], Deny Fauzy[2]
{{yudi_darma@uinsgd.ac.id, ali@uinsgd.ac.id, egiandriana@student.uinsgd.ac.id, ichsan@uinsgd.ac.id, fauzydeny@uinsgd.ac.id }

Ilmu Hadits Department, UIN Sunan Gunung Djati Bandung[1],
Department of Informatics, UIN Sunan Gunung Djati Bandung[2]

Abstract. Al-Qur'an as a guide to one's life implies that Qur'an must also be embraced by Islamic values contained in it, so that values can be a force that motivates and underlies daily activities, and becomes a tool struggle in the field of society or science. It is even more beautiful if that appreciation increases or develops into an effort to improve the study of the interpretation of Qur'an, even trying to memorize it as part of worship. Because memorizing Qur'an for certain people, even though it consists of 30 juz and 114 suras, is not difficult. The problems faced in the current era are the lack of interest in memorizing Qur'an and lack of enthusiasm when told to memorize while muttering together, there are still many who memorize carelessly, while joking at their friends, some even chatting. Similarly, the results of memorization are mainly for memorizing, meaning that there are still many students who have not been able to memorize correctly and fluently. One of the main causes of lack of interest in memorizing the Koran is the method and learning media that are less attractive, seeing this problem is carried out a study in which Al-Quran memorization namely juz 30 is packaged in the form of quizzes with sound input and implemented into a device, smartphone based on android. The method used in this study is Levenshtein Distance. The purpose of this study is to attract interest in memorizing the Koran by packaging it into digital form. Based on the research that has been done, the results show that this application can display according to what is said.

Keywords: Al-Quran, Levenshtein Distance, Android, Smartphone, Application.

1 Introduction

The Holy Qur'an as a guide to one's life implies that, the Qur'an must also be embraced by Islamic values contained in it, so that values can be a force that motivates and underlies daily activities, and becomes struggle tools in the field of society or science. It is even more beautiful if that appreciation increases or develops into an effort to improve the study of the interpretation of the Qur'an [1], even trying to memorize it as part of worship. Because memorizing the Qur'an for certain people, even though it consists of 30 juz and 114 surah, is not difficult. The Holy Qur'an is a basic living guidance for Muslims. It compreses all aspects of human life including Biology, Information Communication and Technology (ICT), Laws, Social, Politics, Business, Economics, Autonomy, and others [2][3].

However, memorizing the Qur'an is not an easy matter but it is also not something that is not possible at this time, in the area where the writer lives there are still many, almost all of the children who after performing the evening prayer have not had time to read the dhikr and prayer,

ICONISTECH 2019, July 11-12, Bandung, Indonesia

DOI 10.4108/eai.11-7-2019.2298059

they immediately ran out of the mosque and immediately stayed in a place not far from the mosque and they immediately played their gags. It would be nice and more useful if after praying the congregation immediately filled in by reading the Qur'an while waiting for the adhan to reverberate. Not only among children as well as what happens with teenagers, it can be counted only a few who after the prayer in the evening sunset immediately read the Qur‘an.

From the explanation above, there are several factors that interest in reading the Qur'an is very low. One of the factors or causes of the decline in interest in reading the Qur'an is the first is technological progress. The use of smartphones in every circle is a necessity that cannot be left behind to continue the daily activities carried out through smartphone media. With this technological advancement they especially children and adolescents prefer to use gadgets such as social media and playing games until they forget time. They prefer to take the time to play gadgets instead of taking the time to at least read the Qur'an which is more beneficial for them, especially now that there are many Holy Qur'an applications that can be downloaded directly on smartphones.

Information and communication technology has developed along with globalization. This requires the development of human resources and education is one of the important things in the development of human resources. For teaching staff to integrate computer technology in learning systems is a challenge, so that learning can be more quality and fun, so it can facilitate students to accept and apply the material presented [4].

The second factor is the decrease in the number of teachers teaching the Koran. Teacher recitation is difficult to find, even among them are sometimes reluctant to teach the Koran or preoccupied with their respective work, in garut regency there is a scarcity of teacher salaries because many clerics / ustadzah are transferring professions to look for other livelihoods due to lack of government attention.

From the problems above, the author thinks hard about how to overcome it, namely by creating learning methods by combining the habits of children and adolescents with smartphones and the interest in reading the Qur'an by building a smartphone application memorizing the Qur'an using the quiz method.

The algorithm used in this study is the Levenshtein Distance algorithm. Levenshtein Distance is a matrix for measuring the number of differences between two strings, the operation needed to convert one string (source string) to another (target string), where an operation involves insertion, deletion, and replacement (substitution) of a single character. Levenshtein Distance algorithm can be applied to an Information Retrieval based search engine to rank query results [5]. This study aims to develop a thesis title / final task search system using the Levenshtein algorithm, which is the subject of research problems. this is in the spelling section used by students in searching for the thesis title. It can be concluded that the Levenshtein algorithm can help overcome the problem of keyword spelling errors by the mechanism of adding, inserting and deleting characters.

Research other to Improve the Effectiveness of Word Search in the Indonesian Dictionary (KBBI)" aims to implement the Levenshtein Distance autocomplete and KBBI applications and to find out effectiveness of its use in the word search feature. The results obtained in this study that the addition of the autocomplete feature is very effective with the value of effectiveness that is 84.615% and with the application of the Levenshtein Distance algorithm implemented in the KBBI application has a percentage of effectiveness of 76.04% [6].

In the research conducted by Peggy and Seng Hansun aimed at optimizing the application for translating Chinese - Indonesian using the Levenshtein Distance algorithm. The results of this study are the application of the Levenshtein Distance algorithm in word search on applications that can translate Chinese into Indonesian successfully, the application of the

Levenshtein Distance algorithm can provide an alternative word for input errors made by the user [7].

In the last few years, computing environments for human learning have rapidly evolved due to the development of information and communication technologies. However, the use of information technology in automatic correction of spelling errors has become increasingly essential. In this context, we have developed a system for correcting spelling errors in the Arabic language based on language models and levenshtein algorithm [8]. The metric distance returned by the Levenshtein algorithm is often the same for multiple solutions in correcting a wrong word. To overcome this limitation we have added a weighting based on language models. This combination has helped us to screen and refine the results obtained in advance by the Levenshtein algorithm, and applied to the errors of Arabic words. The results are encouraging and demonstrate the value of this approach [8].

The Qur'an has a very large number of words so manually searching for a word in the Koran is difficult. There have been many available search applications for verses of the digital Koran based on desktop, mobile and web. But the problem that arises is that the user wants to look for pieces of the Qur'anic verses but does not know the original writing of the verse, only knows the sound. Because of the pronunciation of words, it is often found the pronunciation of the same or similar words but has a different writing. So, a solution is needed to overcome this problem. The development that has been done is phonetic string matching. Phonetic string matching is a string matching algorithm based on the similarity of the sound [9].

From some of the studies above, for the case of the voice recognition Levenshtein Distance algorithm this is very suitable to be used, therefore the author conducted a study entitled "Quran Hifdzil Quiz Application Using the Levenshtein Distance Algorithm".

2 Methodology

The method used in this study is to use the Levenshtein Distance method, because it can be modified when we use this algorithm in it which makes the difference in weight or cost between insertion, deletion, and character replacement. Making this difference in weight can also be further developed by looking at the location of the insertion, deletion or replacement of characters that occur. This distinction aims to get results with more variations in the terms that are the most similar and slightly similar to the terms of the original query.

Research method for data collection techniques in this study are as follows:

1) Literature review is to describe the data obtained from various reference sources related to the problems being studied, to be analyzed, and described in the form of descriptions which are motivated by the existence of concepts and theories put forward in the basic theoretical.
2) Observation, search and collect data, where data has relevance to the title of this final project

a) Levenshtein Distance Algorithm

In information theory and computer science, Levenshtein Distance is a matrix for measuring the number of differences between two strings. Levenshtein Distance of two strings is the minimum number of operations needed to convert one string (source string) to another (target string), where an operation involves insertion, deletion, and substitution of a character

single. Levenshtein Distance is often used in applications to determine how similar or different two strings are, such as checking spelling, or commonly known as spell checkers [10].

Levenshtein distance calculation is a calculation of a matrix used to calculate the number of differences between two strings. Calculation of distance or distance between these two strings is determined by the minimum number of changes to change from string A to string B, which is calculated using the calculation table for levenshtein distance, where the last value in the lower right corner is the final value of the two strings [11][6].

Levenshtein Distance involves the use of a matrix of size (n + 1) x (m + 1), where n and m are the lengths of two strings. Below there is a pseudocode for a Levenshtein Distance function that handles 2 strings, namely the string s with the length m, and the string t with length n. and Levenshtein Distance computing between the two as follows :

```
int LevenshteinDistance(char s[1..m], char t[1..n])
{
    // d is a table with m+1 rows and n+1 columns
    declare int d[0..m, 0..n]
    declare int cost
    for i from 0 to m
      d[i, 0] := i // deletion
    for j from 0 to n
      d[0, j] := j // insertion
       for j from 1 to n
    {
      for i from 1 to m
      {
              if s[i] ≠ t[j] then
           cost := 1
        else
           cost := 0
         d[i, j] := minimum
                    (
                      d[i-1, j] + 1,  // deletion
                      d[i, j-1] + 1,  // insertion
                      d[i-1, j-1] + cost // substitution
                    )
      }
}
```

Tabel 1. Description Pseudocode Program

STEP	DESCRIPTION
	Set n as the length of s Set m as the length of t
1	If n = 0 then distance = m and end If m = 0 then distance = n dan end Build a matrix containing 0..m rows dan 0..n columns
2	Initialize the first line with 0,1,2,…,n Initialize the first column with 0,1,2,…,m
3	Check each character s (from i=1 to i=n)
4	Check each character t (from i=1 to i=m)

5	if s[i] = t[j] then cost = 0. If s[i] ≠ t[j] then cost = 1
6	Set cell d[i,j] from the matrix with a minimum of : - Cell above, added 1 - d[i-1,j] + 1 Cells on the left side, plus 1 - d[i,j-1] + 1 Cells are diagonally top-left, plus cost - d[i-1,j-1] + cost
7	After the iteration step (3,4,5,6) is complete, the Levenshtein Distance is located in cell d[m,n], that is the cell in the lower right corner.

To determine the similarity value can use the following formula similarity :

$$Sim = 1-(\frac{Dist}{Maxlength}) \quad (1)$$

Explanation :

Sim : similarity value

Dist : Distance Levenshtein

MaxLength = Longest string value

The following is the implementation of the Levenshtein Distance method in the voice input process carried out by the user, the report to be matched, namely: ٱللَّهُ ٱلصَّمَدُ

N

		ا	لْ	هُ	صَّ	مَ	دُ	M
	0	1	2	3	4	5	6	
ا	1	0	1	2	3	4	5	
لْ	2	1	0	1	2	3	4	
هُ	3	2	1	0	1	2	3	
صَّ	4	3	2	1	0	1	2	
مَ	5	4	3	2	1	0	1	
دُ	6	5	4	3	2	1	0	

Figure 1. database answer table and answers from users

Looking at the figure above, we compare answers in the database while N we translate the answers spoken by the user. For M [1] and N [1] have the same letter “alif kasrah“, thus we fill 0 because if it has the same letter we have to fill it with the upper left diagonal number. Furthermore for M [2] and N [1] have different letters, we use the condition by summing the numbers that are around M [2] and N [1] with +1 is 2 + 1 = 3, 1 + 1 = 2 and 0 + 1 = 1 and we

take the smallest result to fill in column M [2] N [1]. The number at the bottom right is the distance of the two strings. To determine the similarity value can use the following formula :

$$Sim = 1-(\frac{Dis}{Maxlength}) \quad (2)$$

$Sim = 1-\left(\frac{0}{6}\right) \rightarrow Sim = 1-0 \rightarrow Sim = 1$

b) Scenario the implementation system

Table 2. Definition of Actor Application of Quiz Hifdzil Qur'an

No	Actor	Description
1	User	The actor with this rule has the authority to start games by inputting the name as his account, then the user can play games and get a score and can continue the level of the game if you have answered all the questions in the previous level.

Use case in this app is a process that executes some features that are in the app. Can be seen in table as follows:

Table 3. Definition of Use Case Application for Hifdzil Qur'an Quiz

ID	Requirement	Description
F_01	Playing Quiz	Form to make starting game quiz
F_02	Material	Form for viewing verses 30 material
F_03	Score	Form to see the last score played
F_04	About	Form about the application explanation

Tabel 4. Skenario Use Case Playing Quiz

Action Actor	System
Normal Scenario	
1. Choose The level game playing	
	2. Showing questions about the easy level quiz
3. Answering the question by input sound	
	4. Match the answer with the Levenshthein Distance algorithm
	5. Continue the next questions
	6. Give the final score with at the initial level
7. Continue the level	

Tabel 5. Skenario Use Case Quiz Material

Action Actor	System
Normal Scenario	
1. Start Application	
	2. Showing Login Page
3. Input Username	
	4. Successfully saved username in database
5. Choose Material	
	6. Showing material Quran verses 30

Tabel 6. Skenario Use Case Score Quiz

Action Actor	System
Normal Scenario	
1. Start Application	
	2. Showing Login Page
3. Input Username	
	4. Successfully saved username in database
5. Choose Score History	
	6. The final score appears playing

Tabel 7. Skenario Use Case About Game

Action Actor	System
Normal Scenario	
1. Start Application	
	2. Showing Login Page
3. Input Username	
	4. Successfully saved username in database
5. Choose Score History	
	6. Showing description about game

c) Designing The Quiz Hifdzil Qur'an Application

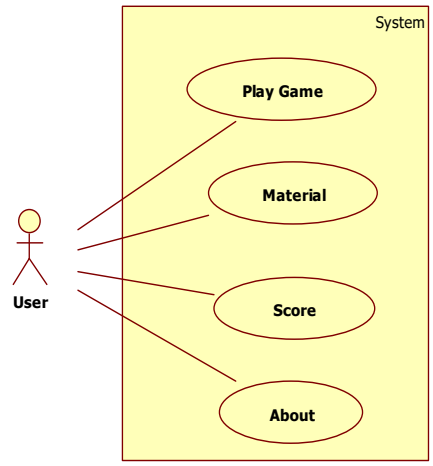

Figure 2. Use Case Diagram Application

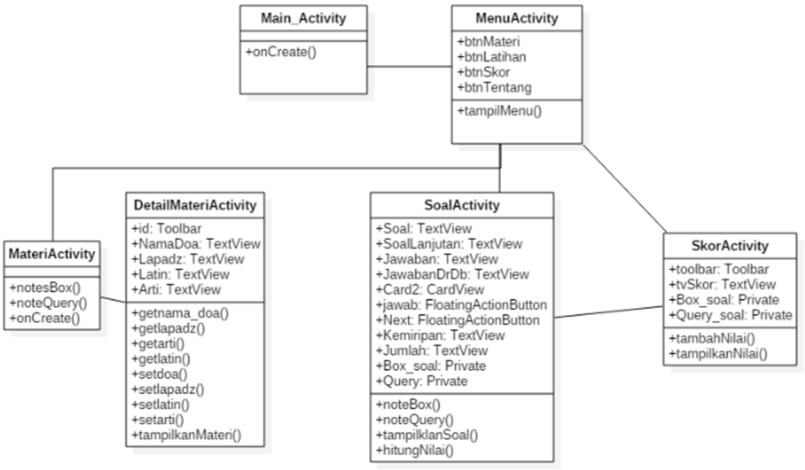

Figure 3. Class Diagram Application

3 Result and Discussion

The following are the results of testing the data obtained after 20 times testing data by answering the questions in the application:

1) Easy level: The calculation value at easy level is quite good, each question gets a calculation value of more than 0.5 value of slippage

2) Medium Level: Calculation value at moderate level is quite good, each question gets a calculation value of more than 0.5 and some even up to 0.8 similarity values
3) Difficult Level: The value of the calculation in the difficult level is quite good, each question gets a number of calculations of more than 0.5 values of slippage

Seeing the results of testing the data above which averaged 0.5 values of flirtation. Some even reached 0.8. This slippage value proved that the Levenshtein Distance method was good enough for the case of voice recognition. After several trials, this tahfidz qur'an test application produces the same appearance of the verse as the application user.

4 Conclusion

With the development of technology, especially in the field of mobile applications, many have made it easier for users to learn learning methods through digital. Then a learning method was made to find out the memorization knowledge of Al-Quran Juz 30 based on mobile applications.

The conducting research on the Hifdzil Quran, users can find out the knowledge of memorizing the Koran especially Juz 30 with the quiz method. With this quiz method you can assess the level of memorization that the user has memorized beforehand (as a knowledge test). By way of inputting answers through voice becomes the attraction of studies that still give answers through text.

Using the Levenshtein Distance method on the Quiz Hifdzil Quran application for this voice recognition case, the sentences are pronounced and the results shown are quite accurate.

Acknowledgements

We would like to say thank you for Research and Publication Centre of UIN Sunan Gunung Djati Bandung that give the full support and fund for this publication research.

References

[1] Shihab M Q 2007 *"Membumikan" Al-Quran: Fungsi dan Peran Wahyu dalam Kehidupan Masyarakat* (Mizan)

[2] Noordin. M F and Malaysia. I I U 2009 *ICT and Islam*

[3] Slamet C, Rahman A, Ramdhani M A and Dharmalaksana W 2016 Clustering the verses of the holy qur'an using K-means algorithm *Asian J. Inf. Technol.* **15** 5159–62

[4] Irfan M and L M R P 2014 Implementasi Computer Based Instruction Model **VIII** 162–76

[5] Arnawa I B K S 2017 Implementasi Algoritma Levenshtein Pada Sistem Pencarian Judul Skripsi / Tugas Akhir *J. Sist. Dan Inform.* **11** 46–53

[6] Ngafidin K and Wibawanto H 2015 Implementasi Fitur Autocomplete dan Algoritma Levenshtein Distance untuk Meningkatkan Efektivitas Pencarian Kata di Kamus Besar Bahasa Indonesia (KBBI) *J. Tek. Elektro* **7** 1–7

[7] Peggy and Hansun S 2015 O ptimasi P encarian K ata pada A plikasi P enerjemah B ahasa M andarin – I ndonesia B erbasis A ndroid dengan A lgoritma L evenshtein D istance *Ultim. Comput.* **VII** 19–23

[8] Lhoussain A S, Hicham G and Abdellah Y 2015 Adaptating the levenshtein distance to contextual spelling correction *Int. J. Comput. Sci. Appl.* **12** 127–33

[9] Arsaningtyas P A, Bijaksana M A and Faraby S Al 2018 Sistem Pencarian Ayat Al-Quran Berdasarkan Kemiripan Ucapan Menggunakan Algoritma Soundex dan Damerau-Levenshtein Distance *J. Linguist. Komputasional* **1** 58–65

[10] Ds E B F, Wirayuda T A B and Suryani A A 2010 PENERAPAN ALGORITMA LEVENSHTEIN DISTANCE PADA DESKTOP SEARCH BERBASIS INFORMATION RETRIEVAL

[11] Ir. Muhammad Aswin, MT., Rachmania Nur Dwitiyastuti., Adharul Muttaqin, ST., MT. 2013 Pengoreksi Kesalahan Ejaan Bahasa Indonesia Menggunakan Metode Levenshtein Distance *J. Mhs. TEUB* **1** 1–6