



Similarity Detection for Hadith of *Fiqh* of Women using Cosine Similarity and Boyer Moore Method

Badruzzaman M Yunus¹, Mohamad Irfan^{2,3}, Wildan Budiawan Zulfikar², Wahyudin Darmalaksana³

^{1,4}Department of Ilmu Alquran and Tafsir, UIN Sunan Gunung Djati Bandung, Indonesia

²Department of ICT, Asia E University, Malaysia

³Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia

E-mail: ¹badruzzaman@uinsgd.ac.id, ²irfan.bahaf@uinsgd.ac.id, ³wildan.b@uinsgd.ac.id,

⁴yudi_darma@uinsgd.ac.id

ABSTRACT

Nowadays, people can get information easily including about fiqh and hadith as a source of Islamic law. The problem is, there are so many articles about jurisprudence whose understanding refers to the laws or rules relating to the hadith whose validity cannot be ascertained. The study aims to determine the degree of similarity between the hadith contained in articles with reliable sources such as books and books. One of the outputs of this study is an application that can determine the similarity of hadith using Cosine Similarity and Boyer Moore by matching strings starting from the right position to the leftmost position and using the cosine similarity method to determine the similarity based on the calculation of the distance between vectors A and B that produce angles cosine x between the two vectors. In the testing phase, the proposed model can run as planned. In one test scenario, the number of keywords tested was 9 cases compared to the categories in the database with an accuracy of 80%. And determine the similarity of two or more objects Using the cosine similarity method with weights The percentage of similarity is proportional to the sample of words entered, which is equal to 36%.

Keywords: Boyer Moore, Cosine Similarity, *Fiqh*, Hadith, Text Mining

1. INTRODUCTION

The rapid development of web applications since the advent of internet technology has greatly helped in the ease and speed of sending, delivering, and receiving information. With the many uses of computers and the use of the internet, people can get information easily including about fiqh for women [1], [2].

Fiqh is one of the fields of science in Islamic law which specifically addresses the legal issues governing various aspects of human life while the *Fiqh* of women is the knowledge of the laws governing the problems of women and discuss funding the sharia rules relating to women [3], [4].

Nowadays, there are many learning websites about women jurisprudence whose understanding refers to the laws or rules relating to women. The case that will be raised in this research is about the similarity of the hadith about the *fiqh* of women which contains specific rules about women to determine the similarity of the hadith on the website with the hadith in the book with clearer sources [5], [6]. There is a hadith on the website that lacks validity, so it takes the media to determine the similarity between the hadith on the website with reliable sources such as books. This is what triggers software developers to create an application that can solve this problem.

Cosine Similarity is one method of determining the similarity between two objects [7], [8]. Cosine Similarity uses two vectors that present two text documents where the cosine angle value of the two vectors is the similarity value of the two text documents. The resulting value limits ranging from 0 to 1 [9], [10]. The composition of words in a text document is a determination of the value obtained from the Cosine Similarity method. The Cosine Similarity method cannot determine the common meaning of each word [11]. Every word that has a different letter component is considered a different word.

This work uses Boyer Moore in string matching to find out the accuracy and accuracy of string matching [12]. Boyer Moore uses sliding and jumping movements [13], [14]. Slide gestures to get suitable characters. The jump motion gives information how many characters must be shifted to match the last character that matches its initial appearance. In string searching it is usually difficult to detect or search for strings (patterns) that are not intact or there may be characters missing in that string [13], [15].

2. METHODOLOGY

2.1 Text Preprocessing

Preprocessing is the initial stage in processing data, preprocessing text is done for the purpose of uniformity and ease of reading [16], [17]. Preprocessing consists of several

stages, at the stage of making this final project only one stage is used, namely case folding. Case folding is a step that changes all the letters in a document into lowercase letters text[18]–[20]. In the end, only the A-Z characters that are accepted by other characters will be eliminated.

```
PUBLIC FUNCTION CASEFOLDING($TEXT)
{
$INPUT = PREG_REPLACE('@[?;:,+=|!~#()0-9]+@', "",
STRTOLOWER($TEXT));
RETURN $INPUT;
}
```

Table 1: Text Preprocessing

Original Document	Processed Document
<p><i>Dalam beberapa adat pernikahan yang ada di Indonesia, kita mungkin pernah atau bahkan sering menjumpai pengantin wanita mencukur habis alisnya karena harus menyesuaikan dengan riasan pengantin di wajahnya. Tidak hanya itu, mencukur alis sampai habis pun sering kali dilakukan oleh banyak wanita yang bekerja di luar rumah untuk mempercantik diri, dengan alasan penampilan adalah penunjang keberhasilan karir mereka. Padahal sesungguhnya perbuatan mencukur alis ini adalah salah satu perbuatan yang dilarang dan diharamkan dalam syariat Islam. Nabi Muhammad shallallahu ,,alaihi wasallam bersabda : Nabi shallallahu,,alaihi wasallam melaknat wanita yang menghilangkan bulu alis dan yang meminta dihilangkan bulu alisnya.” (HR. Abu Dawud, dan terdapat hadits pendukung yang diriwayatkan oleh Al-Bukhari & Muslim) Menghilangkan bulu alis maksudnya adalah mencabut bulu alis atau mencukur bulu alis atau mengerik bulu alis, dan bisa saja dilakukan sendiri baik itu sebagian maupun seluruhnya, dengan alat ataupun dengan tanpa alat. Perbuatan menghilangkan bulu alis ini termasuk perbuatan merubah ciptaan Allah. Karena itu hendaknya setiap wanita menjaga diri dari hal-hal yang diharamkan oleh Allah dan Rasul- Nya. Namun bila seorang wanita menemukan rambut atau bulu yang seharusnya tidak tumbuh pada wajah seorang wanita, seperti kumis dan jenggot, maka ia boleh menghilangkannya karena kumis dan jenggot tadi dapat memberikan mudharat dan memperburuk rupanya. Kodrat seorang wanita adalah ingin selalu tampil cantik, namun tampil cantiknya seorang wanita haruslah dalam koridor syariat. Dimana kecantikan seorang wanita adalah hak suaminya, dan hanya boleh dilihat oleh orang-orang yang menjadi mahramnya. Dan seorang wanita mukminah adalah wanita yang selalu menjaga kehormatan dirinya dan menjaga hak-hak suaminya.</i></p>	<p><i>dalam beberapa adat pernikahan yang ada di indonesia kita mungkin pernah atau bahkan sering menjumpai pengantin wanita mencukur habis alisnya karena harus menyesuaikan dengan riasan pengantin di wajahnya tidak hanya itu mencukur alis sampai habis pun sering kali dilakukan oleh banyak wanita yang bekerja di luar rumah untuk mempercantik diri dengan alasan penampilan adalah penunjang keberhasilan karir mereka padahal sesungguhnya perbuatan mencukur alis ini adalah salah satu perbuatan yang dilarang dan diharamkan dalam syariat islam nabi muhammad shallallahu alaihi wa sallam bersabda nabi shallallahu alaihi wa sallam melaknat wanita yang menghilangkan bulu alis dan yang meminta dihilangkan bulu alisnya hr abu dawud dan terdapat hadits pendukung yang diriwayatkan oleh al bukhari muslim menghilangkan bulu alis maksudnya adalah mencabut bulu alis atau mencukur bulu alis atau mengerik bulu alis dan bisa saja dilakukan sendiri baik itu sebagian maupun seluruhnya dengan alat ataupun dengan tanpa alat perbuatan menghilangkan bulu alis ini termasuk perbuatan merubah ciptaan allah karena itu hendaknya setiap wanita menjaga diri dari hal hal yang diharamkan oleh allah dan rasul nya namun bila seorang wanita menemukan rambut atau bulu yang seharusnya tidak tumbuh pada wajah seorang wanita seperti kumis dan jenggot maka ia boleh menghilangkannya karena kumis dan jenggot tadi dapat memberikan mudharat dan memperburuk rupanya kodrat seorang wanita adalah ingin selalu tampil cantik namun tampil cantiknya seorang wanita haruslah dalam koridor syariat dimana kecantikan seorang wanita adalah hak suaminya dan hanya boleh dilihat oleh orang orang yang menjadi mahramnya dan seorang wanita mukminah adalah wanita yang selalu menjaga kehormatan dirinya dan menjaga hak hak suaminya</i></p>

2.2 Boyer Moore Implementation

The following will be illustrated an example of its application to better understand the search system using the Boyer-Moore algorithm in the application to be made, using the word you want to find (pattern) "HAID" in the text "OLD HAID" [21], [22].

Pattern: "HAID"
Teks: "LAMA HAID"

The first step in matching this string is finding the values for BMBC and BMGS [14], [21], [23]. The value of the results of this preprocessing calculation will be displayed in Table 1.

Table 2: BMBC

BMBC				
Index	0	1	2	3
Pattern	H	A	I	D
BmBc	3	2	1	0

In filling out the values from table 2 this is done with the following conditions: Enumerating starting from the last position of the string to the starting position, starting at the 0th index, note the characters that have been found which in this example the character "D".

Next, back to the previous position, the enumerator value added 1, if the character in this position has never been found, then the shift value is the same as the enumerator value (in this example, the character "I" has never been found before so the shift value is equal to the enumerator value ie 1. Then, move back to the previous position, the character "A", the shift value. And later on getting to the starting position of the string.

Table 3: BMGS

BmGs				
Index	0	1	2	3
Pattern	H	A	I	D
BmGs	4	4	4	1

In filling in the values from the BMGS table, the following conditions are made: BMGS [0] = 4 because the 0th character that H is the left character of the aid segment. There is no more aid segment to the left of the 0h character so it is given a value of 4. BMGS [1] = 4, because the 1st character, a, is the character to the left of the id segment. There is no longer another id segment to the left of the 0h character so it is given a value of 4.

BMGS [2] = 4 because the 2nd character i is the left side of segment d. There is no longer another d segment to the left of the nth character, then a value of 4. BMGS [4] = 1, because the 3rd character d is the leftmost character of the segment, by default it is given a value of 1. The next step is to do a search by aligning the pattern with the text.

Table 4: Iteration Scheme of Boyer Moore

Iteration	0	1	2	3	4	5	6	7	8
1	L	A	M	A		H	A	I	D
	H	A	I	D					
2	L	A	M	A		H	A	I	D
					H	A	I	D	
3	L	A	M	A		H	A	I	D
					H	A	I	D	

Table 4 is an iteration scheme of string searching using the Boyer-Moore algorithm. Iteration 1 shows that the character a in the text does not match the character d in the pattern, so a shift must be made based on the values from the BMBC and BMGS tables by comparing these values and then taking the greatest value for the shift. In the BMBC table a character = 4, character a is given a value of 4 because character a is not in the table, for characters that are not in the table, it is given a value according to the pattern length and the BMGS value. Character d = 1, so the biggest value is 4 to move 4 steps to the right.

Iteration 2 continues from iteration 1 in a pattern that has shifted 4 steps. Matching is performed again, it looks at the iteration of 2 characters i in the text does not match the character d in the pattern. This situation requires a shift back based on the value of BMBC and BMGS by comparing the two values to get the largest value for the shift. In the BMBC table, the character i is worth = 1 and the value of the BMGS character d is = 1, the value taken is 1 so that it moves one step to the right.

Iteration 3 continues from iteration 2 in that point the pattern has shifted 1 step. Matching again is done between the d character in the text with the d character in the pattern matches the result. Continuation matching between the character i in the text with the character i in the pattern matches. Continuing matching between character a in the text and character in the pattern matches the result. Continuation

matching between h characters in the text with h characters in the pattern matches. All characters have found their matches so that the match was completed.

```

PUBLIC FUNCTION RENDER($TEXT, $PATTERN) {
    $PATTERN_TEXT = LTRIM($PATTERN);
    $PATLEN = STRLEN($PATTERN);
    $TEXTLEN = STRLEN($TEXT);
    $TABLE = $THIS->MAKECHARTABLE($PATTERN);
    FOR ($I=$PATLEN-1; $I < $TEXTLEN; ) {
        $T = $I;
        FOR($J=$PATLEN-1;
            $PATTERN[$J]==$TEXT[$I]; $J--, $I--) {
            IF($J == 0) RETURN 1; //MATCH;
        }
        $I = $T;
    }
    IF (ARRAY_KEY_EXISTS($TEXT[$I], $TABLE))
        $I = $I + MAX($TABLE[$TEXT[$I]], 1);
    ELSE
        $I += $PATLEN;
    }
    RETURN 0; //NOT MATCH
}
    
```

2.3 Cosine Similarity Implementation

The principle of similarity shows that the same object will be seen together as a group, this can be determined through shape, color, direction, and size [24], [25]. Following is the pseudo-code of this algorithm:

```

public static function cosinusTokens(array $tokensA,
    array $tokensB) {
    $dotProduct = $normA = $normB = 0;
    $uniqueTokensA = $uniqueTokensB = array();
    $uniqueMergedTokens =
        array_unique(array_merge($tokensA, $tokensB));

    foreach ($tokensA as $key => $token){
        $uniqueTokensA[$token] = count($token);
    }

    foreach ($tokensB as $key => $token){
        $uniqueTokensB[$token] = count($token);
    }

    foreach ($uniqueMergedTokens as $token) {
        $x = isset($uniqueTokensA[$token]) ?
            count($token) : 0;
        $y = isset($uniqueTokensB[$token]) ?
            count($token) : 0;
        $dotProduct += $x * $y;
        $normA += $x;
        $normB += $y;
    }

    $atas = (($normA * $normB) != 0) ? $dotProduct : 0;
    $bawah = sqrt($normA) * sqrt($normB);
    $total = $atas / $bawah;
    return $total;
}
    
```

The dot product is a value that expresses the angle between two vectors. Dot products are scalar values resulting from the operation of two vectors that have the same number of components. If vectors A and B have as many components as n, then the dot product can be calculated with the following formula:

$$(1)$$

Dot products can be calculated by adding up the product of each component in both vectors. If vector A and vector B are 3-dimensional vectors, the dot product calculation is as follows:

$$A \cdot B = A_x \cdot B_x + A_y \cdot B_y + A_z \cdot B_z \quad (2)$$

Calculations to determine the percentage of similarity between documents, then the percentage of similarity is obtained by multiplying the Cosine Similarity value of 100. The following formula for determining the similarity percentage value:

$$\text{Similarity (\%)} = \text{Sim} (A, B) \cdot 100 \quad (3)$$

Cosine Similarity is a method of calculating the distance between vectors A and B which produces an angle of cosine x between the two vectors. The cosine angle value between the two vectors determines the similarity of the two objects being compared where the smallest value is 0 and the largest value is 1. The following is the formula for the Cosine Similarity calculation method:

$$\cos \theta = \frac{a \cdot b}{\sqrt{a^2} \cdot \sqrt{b^2}} \quad (4)$$

Determine the vector values A and B with the number of components n:

$$A = \{a_1, a_2, a_3, \dots, a_n\}$$

$$B = \{b_1, b_2, b_3, \dots, b_n\}$$

Examples:

A: “*Fiqh* related to women’s menstrual blood”

B: “laws related to women’s periodical menstruation”

Table 5: Determination of Vector Values

No	Words	Vector A	Vector B
1	Fiqih	1	0
2	Darah	1	1
3	Wanita	1	1
4	Ilmu	0	1
5	Tentang	0	1
6	Hukum	0	1
7	Mengenai	0	1
8	Yang	0	1
9	Sering	0	1
10	Dialami	0	1
11	Oleh	0	1

Then obtained Vector:

$$A = \{1,1,1,0,0,0,0,0,0,0\}$$

$$B = \{0,1,1,1,1,1,1,1,1,1\}$$

Determine the number of vector values a.b with the number of components as many as n in formula (5).

$$a \cdot b = (a_1 \cdot b_1) + (a_2 \cdot b_2) + \dots + (a_n \cdot b_n) \quad (5)$$

$$a \cdot b = (1 \cdot 0) + (1 \cdot 1) + (1 \cdot 1) + (0 \cdot 1) + (0 \cdot 1) + (0 \cdot 1) + (0 \cdot 1) + (0 \cdot 1) + (0 \cdot 1) + (0 \cdot 1) = 2 \quad (6)$$

$$a \cdot b = 2$$

The value of each vector is entered in formula (5) so that the value of a.b = 2. is obtained. Then run the vectors A and B using formulas (7) and (8), respectively.

$$\|a\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \dots + a_n^2} \quad (7)$$

$$\|b\| = \sqrt{b_1^2 + b_2^2 + b_3^2 + \dots + b_n^2} \quad (8)$$

So we can get the results for A = 1.73 and B = 3.16. After knowing the results of the sum of the values a.b: $\frac{a \cdot b}{\|a\| \|b\|}$ continues with the final calculation using (9).

$$\cos(a, b) = \frac{a \cdot b}{\|a\| \|b\|} \quad (9)$$

$$\begin{aligned} \text{COS (AB)} &= 2 : (1,73 \times 3,16) \\ &= 0,36 \\ &= 0,36 \cdot 100 \% \\ &= 36\% \end{aligned}$$

3. RESULT AND DISCUSSION

3.1 Evaluation Phase

On this work, our proposed model is tested using several cases. Data A is the result of the scraping process and data B is the result of the database.

A = “The question is: what is the law in the process of eyebrow thinning, and the answer is obviously that is forbidden thing to do. According to the prophet Muhammad salallahu alaihi wasalam that he condemns women who lift or whoever asked to deal with their eyebrows and those are considered great sin which were included to *nimsh* category if their eyebrows got cut or lifted as well. The researcher agreed if there is some exception related to the eyebrows thinning if the hair growing is going dense and distract the function of eyes it is allowed to remove some of the hair”.

B = “The Prophet Muhammad SAW condemn women who lift and whoever asked to cut off their eyebrows (H.R Al-Bukhari dan Muslim dan selain mereka)”.

In table 6, it is explained that there are 141 text parts of the document. This section also explains in detail the values and calculations of vectors A and B.

Table 6: Cosine Detail

Parse Document	A	B	A*B	A ²	B ²
pertanyaan	1	0	0	1	0
apa	3	0	0	9	0
hukumnya	2	0	0	4	0
mempertipis	4	0	0	16	0
rambut	6	0	0	36	0
yang	7	2	14	49	4
lebih	1	0	0	1	0
pada	2	0	0	4	0
alis	17	0	0	289	0
mata	10	0	0	100	0
jawaban	1	0	0	1	0
tidak	6	0	0	36	0
boleh	2	0	0	4	0
mengambil	2	0	0	4	0
juga	2	0	0	4	0
mempertipisnya	3	0	0	9	0
karena	2	0	0	4	0
telah	1	0	0	1	0
disabdakan	1	0	0	1	0
oleh	4	0	0	16	0
nabi	1	1	1	1	1
shallallahu	2	0	0	4	0
lsquo	2	0	0	4	0
alaihi	2	0	0	4	0
wa	2	0	0	4	0
sallam	2	0	0	4	0
bahwa	4	0	0	16	0
mencabut	3	0	0	9	0
melaknat	1	1	1	1	1
beliau	1	0	0	1	0
wanita	3	1	3	9	1
matanya	2	0	0	4	0
meminta	1	0	0	1	0
untuk	1	0	0	1	0
dicabut	1	0	0	1	0
para	1	0	0	1	0
ulama	2	0	0	4	0
menjelaskan	1	0	0	1	0
itu	5	0	0	25	0
termasuk	5	0	0	25	0
begitu	1	0	0	1	0
dengan	5	0	0	25	0
merapikan	4	0	0	16	0
jika	4	0	0	16	0
ini	1	0	0	1	0
atau	4	0	0	16	0
mencukur	3	1	3	9	1
cara	3	0	0	9	0
mencabutnya	1	0	0	1	0
maka	3	0	0	9	0
haram	1	0	0	1	0
bahkan	1	0	0	1	0
salah	1	0	0	1	0
satu	1	0	0	1	0
dosa	1	0	0	1	0
besar	1	0	0	1	0
kategori	1	0	0	1	0
nimsh	3	0	0	9	0
dilaknat	1	0	0	1	0
rasulullah	1	0	0	1	0
dipotong	2	0	0	4	0
dicukur	2	1	2	4	1
makruh	2	0	0	4	0
sebagian	2	0	0	4	0
dilarang	1	0	0	1	0
lainnya	1	0	0	1	0
dijadikannya	1	0	0	1	0
berkata	1	0	0	1	0
khusus	1	0	0	1	0

Parse Document	A	B	A*B	A ²	B ²
hal	1	0	0	1	0
saja	1	0	0	1	0
tapi	1	0	0	1	0
umum	1	0	0	1	0
semua	1	0	0	1	0
dakwah	1	0	0	1	0
karya	1	0	0	1	0
syekh	2	0	0	4	0
bin	1	0	0	1	0
baz	1	0	0	1	0
merubah	1	0	0	1	0
allah	1	0	0	1	0
rela	1	0	0	1	0
ada	1	0	0	1	0
di	2	0	0	4	0
wajah	1	0	0	1	0
akan	1	0	0	1	0
tetapi	1	0	0	1	0
kami	2	0	0	4	0
lihat	1	0	0	1	0
selayaknya	1	0	0	1	0
bagi	1	0	0	1	0
sehingga	2	0	0	4	0
mengatakan	1	0	0	1	0
hendaknya	1	0	0	1	0
dia	2	0	0	4	0
melakukannya	1	0	0	1	0
kecuali	1	0	0	1	0
banyak	1	0	0	1	0
sekali	1	0	0	1	0
atas	1	0	0	1	0
sampai	1	0	0	1	0
menutupi	1	0	0	1	0
mengganggu	1	0	0	1	0
penglihatannya	1	0	0	1	0
menghilangkannya	1	0	0	1	0
wallahu	1	0	0	1	0
kumpulan	1	0	0	1	0
sumber	1	0	0	1	0
al	1	0	0	1	0
fatawa	2	0	0	4	0
kitab	1	0	0	1	0
rasa	1	0	0	1	0
ibnu	1	0	0	1	0
utasaimin	1	0	0	1	0

Table 7: Cosine Summarization

No.	Items	A	B
1	Value	954	22
2	Square Root	30.88	4.69

The recapitulation of the calculation is explained in table 7 where the value for A is 954 and B is 22 while the square root value A = 30.88 and B = 4.69.

3.2. Deployment Phase

At this stage Design or system design will be carried out which is the result of transformation of the results of the analysis that has been carried out at the inception stage which will later facilitate the process of making the system. The design is done using the modeling method used, namely UML in accordance with the objectives and functions of each model.

In this process, the user can enter keywords to search and a search results link appears, when the user opens the detail link a data page will appear from the scraping process. In this case there is a case folding process to convert all letters in the document into lowercase letters and result data from poses. scraping and case folding are carried out the process of similarity of data with a database using cosine Similarity so that it can be seen how many similarities between scraping data with data in the database.

This process is using case diagrams illustrate what functionality must be provided by the system whether documented in the use case model that describes the function of the system and the relationships between actors. That illustrates what is done by the system not how a system works. In this work, several actors are involved namely the User and Administrator. Then, users interact with the system in terms of entering keywords and receive complete search results.

Other actors, such as administration, have more interaction, or managing all the resources needed in the model.

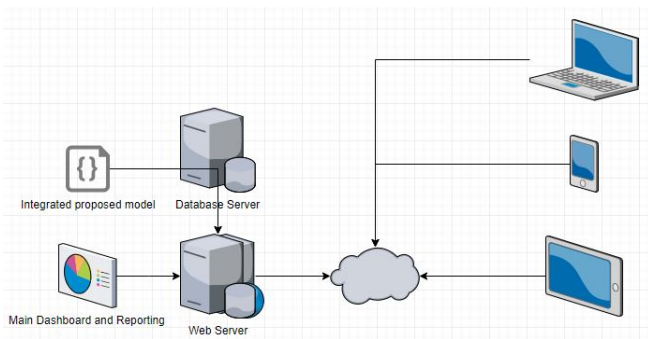


Figure 1: The Architecture of Proposed Model

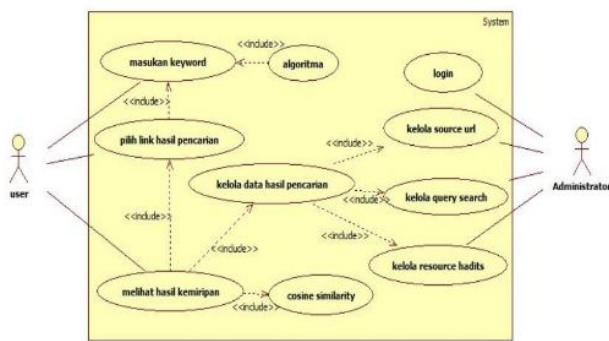


Figure 2: Main Use Case Diagram

Figure 3 explains the main appearance of the application. Users can search for data in the text area that has been provided easily.

Figure 4 illustrates the appearance of a search by entering a keyword and displaying a search result link where in this view there is a Boyer Moore algorithm matching process character from the right of the pattern when the user searches for data based on the input keywords.



Figure 3: The Main interface of Application

Figure 5 explains the scraping details and the similarities between the scraping data and the data in the database. In this view when the user opens the detail will be displayed scraping results in which there is a case folding process and there is also a cosine similarity process to determine the similarity value of data from the search results.

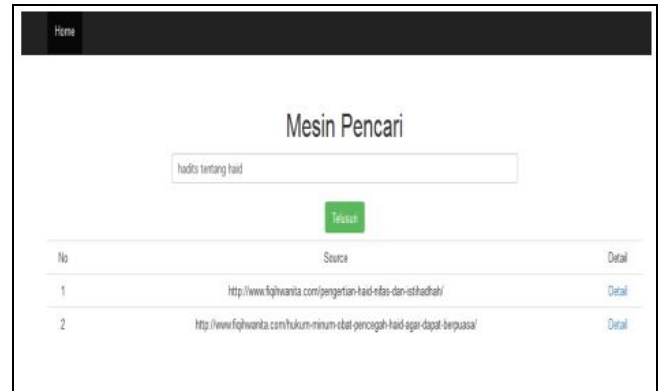


Figure 4: List of Result

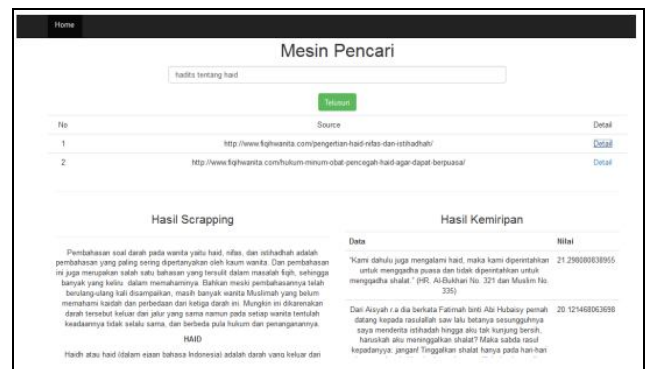


Figure 5: Detail of Result

Black-box testing is carried out to test whether the system developed is following the functional specifications of the system that was previously designed. Black-box testing is done after the system creation process is complete and before the system is distributed to users. Table 8 explains the results of testing using the black-box method or system functionality testing performed in the developer environment.

Table 8: 9 of 99 Test Scenarios by Keyword

No	Keyword	Link	Result	Decision
1	Lamanya waktu haid	http://www.fiqihwanita.com/pengertian-haidnifas-dan-istihadhah/ http://www.fiqihwanita.com/hukum-minumobat-pencegah-haidagar-dapat-berpuasa/	http://www.fiqihwanita.com/pengertian-haidnifas-dan-istihadhah/ http://www.fiqihwanita.com/hukum-minumobat-pencegah-haidagar-dapat-berpuasa/	Correect
2	Puasa	http://www.fiqihwanita.com/hukum-minumobat-pencegah-haidagar-dapat-berpuasa/	http://www.fiqihwanita.com/hukum-minumobat-pencegah-haidagar-dapat-berpuasa/	Correect
3	Hukum mencukur alis	http://www.fiqihwanita.com/hukummencukur-alis/ http://www.fiqihwanita.com/hukummenipiskan-alis-mata/	http://www.fiqihwanita.com/hukummencukur-alis/ http://www.fiqihwanita.com/hukummenipiskan-alis-mata/	Correect
4	Batasan Aurat	http://www.fiqihwanita.com/hukummemakai-wig-di-hadapan-suami/ http://www.fiqihwanita.com/hukummemakai-sepatu-berhak-tinggi-highheels-wedges/	http://www.fiqihwanita.com/hukummemakai-wig-di-hadapan-suami/ http://www.fiqihwanita.com/hukummemakai-sepatu-berhak-tinggi-highheels-wedges/	Correect
5	Shalat	http://www.fiqihwanita.com/hukum-shalatbagi-orang-yangmenjalani-operasi-danpingsan/ http://www.fiqihwanita.com/batakah-shalatjika-kentut-darikemaluan/ http://www.fiqihwanita.com/tata-caramengqadha-shalatyang-terlupa/	http://www.fiqihwanita.com/hukum-shalatbagi-orang-yangmenjalani-operasi-danpingsan/ http://www.fiqihwanita.com/batakah-shalatjika-kentut-darikemaluan/ http://www.fiqihwanita.com/tata-caramengqadha-shalatyang-terlupa/	Correect
6	Puasa bagi orang sakit	http://www.fiqihwanita.com/hukum-minumobat-pencegah-haidagar-dapat-berpuasa/	http://www.fiqihwanita.com/hukum-minumobat-pencegah-haidagar-dapat-berpuasa/	Correect
7	Lamanya nifas	http://www.fiqihwanita.com/pengertian-haidnifas-dan-istihadhah/	http://www.fiqihwanita.com/pengertian-haidnifas-dan-istihadhah/	Correect
8	Hukum KB	http://www.fiqihwanita.com/hukum-kb-dalampandangan-islam/	http://www.fiqihwanita.com/hukum-kb-dalampandangan-islam/	Correect
9	Memakai Kutek	http://www.fiqihwanita.com/wudhunya-wanitayang-kukunyamenggunakan-kutek/	http://www.fiqihwanita.com/wudhunya-wanitayang-kukunyamenggunakan-kutek/	Correect

4. CONCLUSION

Boyer Moore performance algorithm matches the characters from the right of the pattern and the cosine similarity method to find the similarity weights are running well in accordance with the stages. The proposed model can run as planned in the system analysis and design stage. Boyer Moore can be used in search engines on string matching to find out the accuracy and accuracy of string matching, testing is taken from the results of input keywords, with 9 cases compared to the categories in the database with 80% accuracy results. And determine the similarity of two or more objects Using the cosine similarity method with the weight Similarity percentage is proportional to the sample word input, which is equal to 0.36 or 36%. Further work, we suggest adding more comprehensive testing in order to obtain the best result and accuracy value.

REFERENCES

- [1] S. M. Al-Qaththan, *Pengantar Studi Ilmu Hadits*. Jakarta: Pustaka Al-Kautsar, 2013.
[2] H. A. Al-Sanasleh and B. H. Hammo, "Building Domain

Ontology: Experiences in Developing the Prophetic Ontology Form Quran and Hadith," *2017 Int. Conf. New Trends Comput. Sci.*, pp. 223–228, 2017.
<https://doi.org/10.1109/ICTCS.2017.35>

- [3] E. Darwiyanto, G. A. Pratama, and S. Widowati, "Multi words quran and hadith searching based on news using TF-IDF," in *2016 4th International Conference on Information and Communication Technology (ICoICT)*, 2016, pp. 1–6.
<https://doi.org/10.1109/ICoICT.2016.7571915>
[4] P. Hermayanti, A. Budimansyah, and U. T. Lenggana, "Implementasi Metode Scoring System Sebagai Paramater dalam Memahami Kajian Ilmu Tasawuf Berbasis Android," *J. Online Inform.*, vol. 2, no. 2, p. 92, Jan. 2018.
[5] M. Irfan, I. Z. Mutaqin, and R. G. Utomo, "Implementation of Dynamic Time Warping Algorithm on an Android Based Application to Write and Pronounce Hijaiyah Letters," *IEEE CITSM*, 2016.
<https://doi.org/10.1109/CITSM.2016.7577529>
[6] O. Nurdiana, Jumadi, and D. Nursantika, "Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia," *J. Online Inform.*, 2016.

- [7] L. Y. Hu, M. W. Huang, S. W. Ke, and C. F. Tsai, “The distance function effect on k-nearest neighbor classification for medical datasets,” *Springerplus*, vol. 5, no. 1, 2016.
<https://doi.org/10.1186/s40064-016-2941-7>
- [8] M. Alodadi and V. P. Janeja, “Similarity in Patient Support Forums Using TF-IDF and Cosine Similarity Metrics,” in *2015 International Conference on Healthcare Informatics*, 2015, pp. 521–522.
- [9] V. Pratap, S. Tomar, D. Dwivedi, and M. Gwalior, “AN EFFECTUAL APPROACH FOR CALCULATING COSINE SIMILARITY,” *Int. J. Adv. Eng. Res. Dev.*, pp. 1–4, 2015.
- [10] G. Xiaoning, T. De Zhern, S. W. King, T. Y. Fei, and L. H. Shuan, “News Reliability Evaluation using Latent Semantic Analysis,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 16, no. 4, p. 1704, Aug. 2018.
- [11] M. Chahal, “Measuring Similarity between Documents Using TF-IDF Cosine Similarity Function,” *Int. J. Res. Publ. Semin.*, vol. 9, no. 1, pp. 53–57, 2018.
- [12] M. Y. Soleh, “Implementasi Algoritma KMP dan Boyer-Moore dalam Aplikasi Search Engine Sederhana,” *Makal. IF3051 Strateg. Algoritma*.
- [13] Z. Xiong, “A Composite Boyer-Moore algorithm for the string matching problem,” *Parallel Distrib. Comput. Appl. Technol. PDCAT Proc.*, pp. 492–496, 2010.
<https://doi.org/10.1109/PDCAT.2010.58>
- [14] R. Cole, “Tight bounds on the complexity of the Boyer-Moore string matching algorithm,” in *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, 1991, pp. 224–233.
- [15] D. N. K. Manjit Jaiswal, “An Enhanced Boyer-Moore Algorithm for Text String Matching Problem,” *GSTF J. Comput.*, vol. 2, 2012.
- [16] W. B. Zulfikar, M. Irfan, C. N. Alam, and M. Indra, “The comparison of text mining with Naive Bayes classifier, nearest neighbor, and decision tree to detect Indonesian swear words on Twitter,” in *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*, 2017.
- [17] D. Sa’Adillah Maylawati, M. Irfan, and W. Budiawan Zulfikar, “Comparison between BIDE, PrefixSpan, and TRuleGrowth for Mining of Indonesian Text,” *J. Phys. Conf. Ser.*, vol. 801, no. 1, 2017.
- [18] Z. Yao and C. Ze-Wen, “Research on the construction and filter method of stop-word list in text preprocessing,” *Proc. - 4th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2011*, vol. 1, pp. 217–221, 2011.
<https://doi.org/10.1109/ICICTA.2011.64>
- [19] Y. Bao, G. Yang, and W. Jin, “Evaluation of stop word list in Mongolian,” *J. Inf. Comput. Sci.*, vol. 6, no. 3, pp. 1139–1145, 2009.
- [20] N. A. Rakhmawati, A. A. Firmansyah, P. M. Effendi, R. Abdillah, and T. A. Cahyono, “Auto Halal Detection Products Based on Euclidian Distance and Cosine Similarity,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, pp. 1706–1711, 2018.
- [21] H. M. Mahmoud, R. T. Smythe, and M. Régnier, “Analysis of Boyer-Moore-Horspool string-matching heuristic,” *Random Struct. Algorithms*, vol. 10, no. 1–2, pp. 169–186, 2000.
- [22] J. Tarhio and E. Ukkonen, “Approximate Boyer-Moore string matching,” *SIAM J. Comput.*, vol. 22, no. 2, pp. 243–260, 1993.
- [23] Zhu Rui Feng and T. Takaoka, “ON IMPROVING THE AVERAGE CASE OF THE BOYER-MOORE STRING MATCHING ALGORITHM.,” *J. Inf. Process.*, vol. 10, no. 3, pp. 173–177, 1987.
- [24] H. O. Alani and S. Saad, “Schema Matching for Large-Scale Data Based on Ontology Clustering Method,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, pp. 1790–1797, 2017.
<https://doi.org/10.18517/ijaseit.7.5.2133>
- [25] F. A. Bohani, S. N. H. Sheikh Abdullah, and K. Omar, “Cosine Harmony Search (CHS) for Static Optimization,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, pp. 1753–1761, 2018.
<https://doi.org/10.18517/ijaseit.8.4-2.6798>