Comparative Analysis between Rabin Karp Algorithm, Winnowing, and Turnitin Applications for Detecting Plagiarized Words

Ari Kurniawan Saputra [1], Kurnia Muludi[2] , Taqwan Thamrin [3]
[1]*Faculty of Computer Science, Bandar Lampung University, Bandar Lampung City, Lampung Province, Indonesia*
[2]*Department of Computer Science, Lampung University, Bandar Lampung City, Lampung Province, Indonesia*
[3]*Faculty of Computer Science, Bandar Lampung University, Bandar Lampung City, Lampung Province, Indonesia*
ari.kurniawan@ubl.ac.id, kmuludi@yahoo.com, taqwanthamrin@ubl.ac.id

Abstract: Plagiarism in works such as scientific writing said to violate the law because it has stolen the copyrights of other people's works. To prevent plagiarism in the case of scientific writing like this, it can been done by detecting every word or sentence in it using a string or word-matching algorithm in the written text. This study will compare the results of the accuracy between the Rabin Karp Algorithm and the Winnowing Algorithm, the accuracy of the two algorithms will been compared with the results of the similarity value of the text with the Turnitin application. Based on the research of Comparative Analysis of Rabin Karp Algorithm, Winnowing Algorithm and Turnitin Application in calculating the similarity of text on plagiarism detection which is tested using the Hypothesis T-test. Several conclusions were obtained, namely: (1) T-test Hypothesis Testing between testing the average value of the Rabin Karp Algorithm and Winnowing Algorithm and the hypothesis formulation $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ generated value $T_{hitung} = 4{,}222$ and T_{tabel} value = 2,021 so that the decision of the T-test Hypothesis Test was that H_1 was accepted because $T_{hitung} > T_{tabel}$, (2) T-test Hypothesis Test between testing the average value of the Rabin Karp Algorithm and Turnitin Application with the hypothesis formulation$H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ generated value $T_{hitung} = 12{,}505$ and T_{tabel} value = 2,021 so that the decision of the T-test Hypothesis Test was that H_1 was accepted because $T_{hitung} > T_{tabel}$, (3) T-test Hypothesis Testing between testing the average value of Winnowing Algorithm and Turnitin Application with hypothesis formulation $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ generated value $T_{hitung} = 4{,}511$ and T_{tabel} value= 2,021 so that the decision of the T-test Hypothesis Test was that H1 was accepted because $T_{hitung} > T_{tabel}$. It was stated that the Rabin Karp Algorithm had a higher test value weight than the Winnowing Algorithm and the Turnitin Application.

Keywords: Plagiarism, Turnitin, Rabin Karp Algorithm, Winnowing Algorithm

1. INTRODUCTION

The problem of plagiarism in currently digital era was able to be done easily. This plagiarism practice often occurred, especially among academics. The occurrence of plagiarism was due to the habit of an academic who wanted to be fast in completing his assignments so that the possibility of looking for references using digital technology by utilizing the copy-paste technique (copy-paste) an academician was easy to do a plagiarism. The classification of the proportion or percentage of plagiarized words, sentences, and paragraphs was: (1) mild plagiarism - the proportion or percentage of plagiarized words, sentences or sentences was not more than 30 percent (<30%), (2) moderate plagiarism - the proportion or percentage of plagiarized words, sentences or paragraphs was between 30-70 percent, and (3) heavy plagiarism - the proportion or percentage of plagiarized words, sentences, paragraphs was more than 70 percent (> 70%) [11]. This study was comparing the results of the accuracy between the Rabin Karp Algorithm and the Winnowing Algorithm. The accuracy of the two algorithms was compared with the results of the similarity value of the text with the Turnitin application.

The Rabin Karp algorithm was chosen because this algorithm was able to detect text similarities in documents by using the hash method in searching for a word. This theory was rarely used to search for singular words but it was quite important and very effective when used for plural searches [1]. Winnowing algorithm was chosen because of the hash values of each k-gram was calculated to find the hash value then the rolling

hash function was used. A window was formed from these hash values. In each window, the minimum hash value was selected. If there was more than one hash with the minimum value, the rightmost hash value was selected. All selected hash values were stored and used as fingerprints of a document. This fingerprint was used as a basis for comparing the similarities between the text that had been entered. [3]

2. METHOD

2.1 Rabin Karp Algorithm

According to [1] the Rabin Karp algorithm was invented by Michael O. Rabin and Richard M. Karp. This algorithm used the hash method in finding a word. This theory was rarely used to search for singular words but it was quite important and very effective when it was for plural searches. Rabin Karp represented each character as a decimal digit *(digit radix-d)* $\sum = \{0, 1, 2, 3, \ldots, d\}$, where$d = |\sum|$ so that we reached the input string k in succession as a representative of the length of *k* decimal. Character*string* 31415 corresponded to the number of decimals 31,415. Then the p pattern was hashed into a decimal value and the string was represented by the sum of the digits using Horner's rule, for example:
{ A, B, C, ..., Z } → { 0, 1, 2, ..., 26 }

• *BAN → 1 + 0 + 13 = 14*

• *CARD → 2 + 0 + 17 + 3 = 22*

For long patterns and large text, this algorithm used the mod operation, after applying the mod q operation, the value was smaller than q, for example:
BAN = 1 + 0 + 13 = 14 = 14 mod 13 = 1 = BAN → 1

• *CARD= 2 + 0 + 17 + 3 = 22 = 22 mod 13 = 9 = CARD → 9*

Meanwhile, the pseudo-code and mathematical formula used was as follows

```
RABIN-KARP-MATCHER (T, P, d,q)
n = T.length
m = P.length
h = d^(m-1) mod q
p = 0
t0 = 0
for I = 1 to m
  p = (dp + P[I] mod q
  t0 = (dt0 + T[I]) mod q
for s = 0 to n – m
  if p == ts
            if P[1 .. m] == T[s + 1 .. s + m]
                      print “Pattern occurs with shift” s
  if s<n – m
            ts+1 = (d(ts – T[s + 1] h) + T[s + m + 1]) mod q
```

The Rabin Karp algorithm steps were as follows:

1. Parsing. It was a term that needed the preprocessing process cut into pieces per character. Truncation per character using the k-Grams method. The method of k-grams worked by taking pieces of the letter k number of characters from a word which was continuously read from the source text to the end of the document. example k-grams where k = 4:
   Text: evenifnone

   Result 4-grams of text: even | veni | enif | nifn | ifno | fnon | none

2. *Hashing*. It converted a chunk of the number k into a hash value. It usually converted to ASCII. Equation 1 was the formula for Rabin Karp hashing.

$$Hash = T[i]b^{m-1} + T[i]b^{m-2} + \ldots + T[i + m-1] mod\ q.. \quad (1)$$

Where,
T[i] = ASCII value of the i-th index letter
b = *radix* decimal (base 10 number)
m = length of text
q = number of patterns

3. Improving the performance of Rabin Karp with Equation 2, providing a solution for not only comparing the remainder of the quotient, but compare the results for him too.

$$REM\ (n1/q) = REM\ (n2/q)\ and\ QUOTIENT\ (n1/q) = QUOTIENT\ (n2/q)\ \ldots \quad (2)$$

Value Measurement, according to S. Kosinov (2001), parsing was done with the k-gram approach, to measure the Similarity value using *Dice's Similarity Coefficient*, the mathematical formula was shown in equation 3.

$$S = \frac{K\ X\ C}{A+B} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Where S was the value of Similarity, A and B were the sum of the set of k-grams in text 1 and text 2. C was the number of equal k-grams of text being compared. The correct Similarity measure not only improved the quality of information selection but it also helped reduce processing time and costs.

2.2 Winnowing Algorithm

According to [18] the Winnowing algorithm was an algorithm used to carry out the process of checking word similarity (document fingerprinting) to identify plagiarism. To detect documents using the Winnowing algorithm, several formulas were needed, among others:

1. Searching for the first *hash*

$$H_{(C1 \ldots CK)} = c1 * b^{(k-1)} + c2 *\ b^{(k-2)} + \cdots + C^{(k-1)} * b + ck \quad (4)$$

2. Searching for the second *hash*

$$H_{(c2 \ldots ck+1)} = \left(H_{(c1 \ldots ck)} - c1 * b^{(k-1)}\right) * b + c^{(k+1)} \quad (5)$$

3. Measuring the similarity value using *Jaccard Coefficient.*

$$\text{Similarity}\left(d_i d_j\right) = \frac{|w(d_i) \cap (d_j)|}{|w(d_i) \cup (d_j)|} \quad (6)$$

In detecting plagiarism, there were fundamental requirements that must be met by a plagiarism algorithm, e.g.: *Whitespace Insensitivity*, *Noise Suppression*, and *Position Independence*. The meaning of *Whitespace Insensitivity wa*s a match processes against text files and should not be affected by spaces, fonts (capital or normal), punctuation and so on. *Noise suppression* was to avoid finding matches with word lengths that were too small or less relevant, e.g.: *'the'*. The length of the word that was suspected to be plagiarized must be sufficient to prove that the words had been plagiarized and were not a commonly used word. And *Position Independence*, which meant finding a match / similarity did not have to depend on the position of the words. Although it was not in the same position, matching must also be done.

2.3 Turnitin Application

According to [16]Turnitin.com (2016b) categorized 10 types of plagiarism originating from a survey of nearly 900 secondary and high school education instructors worldwide. The ten types of plagiarism were:

1. *Clone*, clone or write the exact same writing of another person word for word.
2. CTRL+C, writing

3. *Find Replace*, replacing words and key phrases from other sources while maintaining the main content.
4. *Remix*, paraphrasing from various sources and writing them together (without source).
5. *Recycle*, borrowing the other people's previous writings without including the source.
6. *Hybrid*, combining other people's writings by inserting one other article and still including one source without performing citation procedures.
7. *Mashup*, mixing various writings / material from various sources.
8. 404 *Error*, using someone else's writing with an inaccurate source or even write a fabricated source.
9. *Aggregator*, doing a citation appropriately (change other people's writing) but almost none of the sources were original (making it up).
10. *Re-Tweet*, doing citation accurately, but it was too close to the original wording or structure.

2.4 Dataset (Corpus)

The dataset (corpus) used in this study used two algorithms, e.g., Rabin Karp Algorithm, Winnowing Algorithm, and Turnitin Application. The dataset used in this research was 30 paper datasets.The documents that were used as the dataset were compared with the tested data documents which calculated the similarity of the text using the Rabin Karp Algorithm, Winnowing and the Turnitin Application, then to obtain the results of the percentage of text similarities used in the plagiarism detection process using these two algorithms.

2.5 Measurement Stage

The measurement stage in this study used the t-test average value comparison method. This test was a comparative test to assess the difference between a certain value and the average population group. The hypothesis carried the use of the term Zero Hypothesis (H_0). Rejection of H_0 results in acceptance of an Alternative Hypothesis denoted by H_1. The zero hypothesis, regarding a population, the parameter must be pronounced in such a way as to state with certainty a value for that parameter whereas the alternative hypothesis allowed several possible values. So if H_0 stated the Zero Hypothesis that $p = 0.5$ for a bionomic population, then the alternative hypothesis H_1 was $p > 0.5$, $p < 0.5$ or $p \neq 0.5$ [20]. The average test value of the two algorithms for each corpus was calculated through the standard deviation as follows:

$$S = \sqrt{\frac{n\sum_i^n = 1\, x_1^2 - (\sum_i^n = 1\, x_1)^2}{n(n-1)}} \quad (7)$$

Information:
s^2 = variant
s = standard deviation (standard deviation)
xi = value$x$ to-i
$\bar{x}$ = average
n = sample size

The next step was conducting a comparative test between the average value of the two algorithms using the t-test hypothesis as follows:

$$t - test = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{s_1^2}{N_{1-1}}\right) + \left(\frac{s_2^2}{N_{2-1}}\right)}} \text{ with } SD_1^2 = \left[\frac{\sum X_1^2}{\mathrm{N}_1} - (X_1)^2\right] \quad (8)$$

Information:
$\bar{X}_1$ = Average sample distribution 1
$\bar{X}_2$ = Average sample distribution 2
SD_1^2 = The value of variance in the sample distribution 1
SD_2^2 = The value of variance in the sample distribution 2
N_1 = Number of papers in the sample 1
N_2 = Number of papers in the sample l 2

4. RESULT AND DISCUSSION

3.1 Result

To be able to find out the highest text similarity percentage weight value between the Rabin Karp Algorithm, Winnowing Algorithm, and Turnitin Application was through the reference value for the percentage of text similarities. The highest weight test of the average plagiarism test value for each algorithm was used to test the T-test Hypothesis which was wrongly one statistic used to test the truth or falsity of the hypothesis which stated that between the two sample means taken randomly from the same population. There was no significant difference. Hypothesis testing T-test was carried out between the Rabin Karp Algorithm and the Winnowing Algorithm, the Rabin Karp Algorithm and the Turnitin Application, the Winnowing Algorithm and the Turnitin Application. The test results are shown in Table 1: Table of Plagiarism Test Results below.

Table 1: Table of Result Test Plagiarism

No	File Testing	Plagiarism Test Average Value		
		Rabin Karp Algorithm	Winnowing Algorithm	Turnitin application
1	Paper 01	35.18%	37.68%	26.71%
2	Paper 02	46.64%	33.49%	20.23%
3	Paper 03	48.04%	40.08%	28.93%
4	Paper 04	48.91%	30.12%	28.28%
5	Paper 05	46.10%	38.03%	35.13%
6	Paper 06	40.79%	36.75%	29.00%
7	Paper 07	40.30%	11.07%	7.62%
8	Paper 08	40.25%	36.96%	34.50%
9	Paper 09	22.63%	21.54%	15.55%
10	Paper 10	39.89%	22.67%	20.00%
11	Paper 11	48.62%	45.58%	32.91%
12	Paper 12	45.03%	36.85%	32.44%
13	Paper 13	38.18%	34.98%	26.90%
14	Paper 14	46.61%	39.93%	26.11%
15	Paper 15	45.99%	41.11%	30.13%
16	Paper 16	39.86%	33.56%	25.00%
17	Paper 17	39.88%	32.75%	26.00%
18	Paper 18	42.42%	34.66%	25.00%
19	Paper 19	30.89%	17.55%	24.00%
20	Paper 20	47.31%	27.88%	25.04%
21	Paper 21	40.67%	36.89%	30.68%
22	Paper 22	42.72%	39.47%	36.00%
23	Paper 23	34.03%	47.86%	39.06%
24	Paper 24	43.92%	36.61%	30.00%
25	Paper 25	48.09%	43.16%	30.00%
26	Paper 26	43.73%	38.53%	21.05%
27	Paper 27	44.61%	37.91%	24.00%
28	Paper 28	37.33%	42.09%	23.40%
29	Paper 29	37.75%	35.78%	18.00%
30	Paper 30	35.87%	42.94%	26.00%
Standar Deviasi		0.06%	8.12%	6.49%
Rata-rata		41.41%	35.15%	26.59%

3.1.1 Hypothesis Test of Rabin Karp Algorithm and Winnowing Algorithm

Known:

$n_1 = 30$

$n_2 = 30$

$\bar{X}_1 = 41{,}41$

$\bar{X}_2 = 35{,}15$

$S_1 = 0{,}06$

$S_2 = 8{,}12$

Information:

n = Sample value (n) or test value
$\bar{X}$ = The average value of the testing
S = The standard deviation value of the testing
Step:

1. Hypothesis Formulation

   $H_0 : \mu_1 = \mu_2$

   $H_1 : \mu_1 > \mu_2$

2. Test Level

   $\alpha = 5\ \% = 0.05$

   df *(degree of freedom)* $= n_1 + n_2$

   $df = 30 + 30 - 2 = 58$

3. T_{tabel} Value

   $T_{tabel} = T\ \alpha / 2\ (df)$

   $T_{tabel} = 0.025\ (58) = 2{,}002$

4. Calculation

$$S_p^2 = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2} = \frac{(30-1)0{,}06^2+(30-1)8{,}12^2}{30+30-2} = 32{,}969$$

$$S_p = \sqrt{32{,}969} = 5{,}742 \qquad (9)$$

$$T_{hitung} = \frac{\bar{X}_1-\bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{41{,}41-35{,}15}{5{,}742\sqrt{\frac{1}{30}+\frac{1}{30}}} = 4{,}222$$

5. Decisions

   H_1 was received because $T_{hitung} > T_{tabel}$ then the test results of the average value of Rabin Karp Algortima were higher than the average value of the Winnowing Algorithm.

3.1.2 Hypothesis Test of Rabin Karp Algorithm and Turnitin Application

Known:

$n_1 = 30$

$n_2 = 30$

$\bar{X}_1$ = 41,41

$\bar{X}_3$ = 26,59

S_1 = 0,06

S_3 = 6,49

Information:
n = Sample value (n) or test value
$\bar{X}$ = The average value of the testing
S = The standard deviation value of the testing

Step:

1. Hypothesis Formulation

   H_0 : $\mu_1 = \mu_2$

   H_1 : $\mu_1 > \mu_2$

2. Test Level

   α = 5 % = 0.05

   df *(degree of freedom)* = $n_1 + n_3$

   df = 30 + 30 – 2 = 58

3. T_{tabel} Value

   $T_{tabel} = T\ \alpha / 2\ (df)$

   T_{tabel} =0.025 (58) = 2,002

4. Calculation

$$S_p^2 = \frac{(n_1-1)S_1+(n_3-1)S_3}{n_1+n_3-2} = \frac{(30-1)0,06^2+(30-1)6,49^2}{30+30-2} = 21,062$$

$$S_p = \sqrt{21,062} = 4,59 \quad (10)$$

$$T_{hitung} = \frac{\bar{X}_1-\bar{X}_3}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_3}}} = = \frac{41,41-26,59}{4,59\sqrt{\frac{1}{30}+\frac{1}{30}}} = 12,505$$

5. Decisions

   H_1 was received because$T_{hitung} > T_{tabel}$ then the test results of the average value of Rabin Karp Algortima were higher than the average value of the Turnitin Application.

3.1.3 Hypothesis Testing Winnowing Algorithm T-test and Turnitin Application

Known:

n_2 = 30

n_3 = 30

$\bar{X}_2$ = 35,15

$\bar{X}_3$ = 26,59

S_2 = 8,12

S_3 = 6,49

Information:

n = Sample value (n) or test value
$\bar{X}$ = The average value of the testing
S = The standard deviation value of the testing

Step:

1. Hypothesis Formulation

   $H_0 : \mu_2 = \mu_3$

   $H_1 : \mu_2 > \mu_3$

2. Test Level

   $\alpha = 5\ \% = 0.05$

   df *(degree of freedom)* $= n_1 + n_3$

   df = 30 + 30 – 2 = 58

3. T_{tabel} Level

   $T_{tabel} = T\ \alpha / 2\ (df)$

   $T_{tabel} = 0.025\ (58) = 2{,}002$

4. Calculation

$$S_p^2 = \frac{(n_2-1)S_2+(n_3-1)S_3}{n_2+n_3-2} = \frac{(30-1)8{,}12^2+(30-1)6{,}49^2}{30+30-2} = 54{,}027$$

$$S_p = \sqrt{54{,}027} = 7{,}35 \qquad (11)$$

$$T_{hitung} = \frac{\bar{X}_2-\bar{X}_3}{S_p\sqrt{\frac{1}{n_2}+\frac{1}{n_3}}} = = \frac{35{,}15-26{,}59}{7{,}35\sqrt{\frac{1}{30}+\frac{1}{30}}} = 4{,}511$$

5. Decisions

   H_1 was received because $T_{hitung} > T_{tabel}$ then the test results of the average value of Algorithm Winnowing were higher than the average value of the Turnitin Application

4.2 Discussion

In the T-test Hypothesis testing, the highest value weight between the average value testing of the Rabin Karp Algorithm and the Winnowing Algorithm and the Turnitin Application used as a reference value for text similarity was shown in the following Table 2: Table of Hypothesis Testing Results T-test.

Table 2: Table of Hypothesis Testing Results T-test

Algorithm / Application	Hypothesis Formulation	Decision on Hypothesis Testing T-test
*Rabin Karp*Algorithm–*Winnowing*Algorithm	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 > \mu_2$	H_1 was accepted because $T_{hitung} > T_{tabel}$ then the test results of the average value of Rabin Karp Algortima were higher than the average value of the Winnowing Algorithm
*Rabin Karp*Algorithm–*Turnitin Application*	$H_0 : \mu_1 = \mu_3$ $H_1 : \mu_1 > \mu_3$	H_1 was accepted because $T_{hitung} > T_{tabel}$ then the test results of the average value of Rabin Karp Algortima were higher than the average value of the Turnitin Application

Algorithm Winnowing - *Turnitin Application*	$H_0 : \mu_1 = \mu_3$ $H_1 : \mu_1 > \mu_3$	H1 was accepted because Thitung > Ttabel then the test results of the average value of Algortima Winnowing were higher than the average value of the Turnitin Application

From the results of the Hypothesis T-test table above, it showed that the Rabin Karp Algorithm resulted the highest average value test between the Winnowing Algorithm and the Turnitin Application.

5. CONCLUSION

According to the Comparative Analysis of Rabin Karp Algorithm, Winnowing Algorithm and Turnitin Application. There were several conclusions e.g.:

1. T-test Hypothesis Testing between testing the average value of the Rabin Karp Algorithm and Winnowing Algorithm and the hypothesis formulation $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ generated value T_{hitung} = 4,222 and T_{tabel} value = 2,021 so that the decision of the T-test Hypothesis Test was that H1 was accepted because $T_{hitung} > T_{tabel}$
2. T-test Hypothesis Test between testing the average value of the Rabin Karp Algorithm and Turnitin Application with the hypothesis formulation $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ generated value T_{hitung} = 12,505 and T_{tabel} value = 2,021 so that the decision of the T-test Hypothesis Test was that H1 was accepted because $T_{hitung} > T_{tabel}$
3. T-test Hypothesis Testing between testing the average value of Winnowing Algorithm and Turnitin Application with hypothesis formulation $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ generated value T_{hitung} = 4,511 and T_{tabel} value= 2,021 so that the decision of the T-test Hypothesis Test was that H1 was accepted because $T_{hitung} > T_{tabel}$

It was stated that the Rabin Karp Algorithm had a higher test value weight than the Winnowing Algorithm and the Turnitin Application.

REFERENCES

[1]. Adi, S. (2018). Penerapan Algoritma Rabin Karp Untuk Mendeteksi Kemiripan Judul Skripsi. *Jurnal Mantik Penuasa*, *22*(1), 125–130.

[2]. Afdhal; Taufan Chalis; Tauiq A. Gani. (2014). Analisa Perbandingan Aplikasi Pendeteksi Plagiat Terhadap Karya Ilmiah. *Seminar Nasional Dan Expo Teknik Elektro*, *May*, 193–199. https://www.researchgate.net/publication/275657950%0AAnalisa

[3]. Alamsyah, N. (2017). Perbandingan Algoritma Winnowing Dengan Algoritma Rabin Karp Untuk Mendeteksi Plagiarisme Pada Kemiripan Teks Judul Skripsi. *Technologia: Jurnal Ilmiah*, *8*(3), 124. https://doi.org/10.31602/tji.v8i3.1116

[4]. Baker, R. K., & Adams, M. (2008). An Evaluation Of The Effectiveness Of Turnitin.Com As A Tool For Reducing Plagiarism In Graduate Student Term Papers. *College Teaching Methods & Styles Journal-September*, *4*(9), 1. http://www.turnitin.com/static/plagiarism.html

[5]. Halgamuge, M. N. (2017). The use and analysis of anti-plagiarism software: Turnitin tool for formative assessment and feedback. *Computer Applications in Engineering Education*, *25*(6), 895–909. https://doi.org/10.1002/cae.21842

[6]. Hasan, E. G., Wicaksana, A., & Hansun, S. (2018). The Implementation of Winnowing Algorithm for Plagiarism Detection in Moodle-based E-learning. *Proceedings - 17th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2018*, 321–325. https://doi.org/10.1109/ICIS.2018.8466429

[7]. Kharisman, O., Susanto, B., & Suwarno, S. (2013). Implementasi Algoritma Winnowing Untuk Mendeteksi Kemiripan Pada Dokuken Teks. In *Jurnal Informatika* (Vol. 9, Issue 1). https://doi.org/10.21460/inf.2013.91.141

[8]. Kurniawati, A., & Wicaksana, I. W. S. (2008). Perbandingan Pendekatan Deteksi Plagiarism Dokumen Dalam Bahasa Inggris. *Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen (KOMMIT 2008)*, *Kommit*, 284–291. https://doi.org/10.1038/473419b

[9]. Leman, D., Rahman, M., Ikorasaki, F., Riza, B. S., & Akbbar, M. B. (2019). Rabin Karp and Winnowing Algorithm for Statistics of Text Document Plagiarism Detection. *2019 7th International Conference on Cyber and IT Service Management, CITSM 2019*. https://doi.org/10.1109/CITSM47753.2019.8965422

[10]. Noprisson, H., Susilo, B., Studi, P., Informatika, T., Teknik, F., & Bengkulu, U. (2013). *IMPLEMENTASI ALGORITMA RABIN-KARP UNTUK MENENTUKAN KETERKAITAN ANTAR PUBLIKASI PENELITIAN DOSEN TAHUN 2013 (Studi Kasus : Website Lembaga Penelitian Universitas Bengkulu)*. *9*, 1–15.

[11]. Purba, A. H., & Situmorang, Z. (2017). Analisis Perbandingan Algoritma Rabin-Karp Dan Levenshtein Distance Dalam Menghitung Kemiripan Teks. *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, *02*, 24–32.

[12]. Purba, L. J., & Sitorus, L. (2018). Perancangan Aplikasi Untuk Menghitung Persentase Kemiripan Proposal Dan Isi Skripsi Dengan Algoritma Rabin-Karp. *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, *3*(1), 17–25.

[13]. Rachmaindah, A., & Rohmiyati, Y. (2018). Analisis Tingkat Plagiarisme Karya Ilmiah Jurnal Mahasiswa Ilmu Perpustakaan Undip Tahun 2015-2016 Dengan Software Turnitin. *Jurnal Ilmu Perpustakaan*, *7*(No. 3), 11–20. https://ejournal3.undip.ac.id/index.php/jip/article/view/22914

[14]. Shadiqi, M. A. (2019). Memahami dan Mencegah Perilaku Plagiarisme dalam Menulis Karya Ilmiah. *Buletin Psikologi*, *27*(1), 30. https://doi.org/10.22146/buletinpsikologi.43058

[15]. Sibarani, L., Magdalena, M., & Dharma, A. (2019). Analisa Perbandingan Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing Dan Algoritma Rabin Karp. *REMIK (Riset Dan E-Jurnal Manajemen Informatika Komputer)*, *4*(1), 69. https://doi.org/10.33395/remik.v4i1.10174

[16]. Sinaga, S. Y. (2018). Penggunaan Aplikasi Turnitin Sebagai Sarana Cek Plagiarisme dalam Layanan Perpustakaan Universitas Ukrida. *BIBLIOTIKA : Jurnal Kajian Perpustakaan Dan Informasi*, *2*(2), 123–131. https://doi.org/10.17977/um008v2i22018p123

[17]. Sugiono, S., Herwin, H., Hamdani, H., & Erlin, E. (2018). Aplikasi Pendeteksi Tingkat Kesamaan Dokumen Teks: Algoritma Rabin Karp Vs. Winnowing. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, *9*(1), 82–93. https://doi.org/10.31849/digitalzone.v9i1.1242

[18]. Ulfa, N. F., Mustikasari, M., & Bastian, I. (2016). Pendeteksian tingkat similaritas dokumen berbasis web menggunakan algoritma winnowing. *Konferensi Nasional Teknologi Informasi Dan Komunikasi (KNASTIK)*, *November*, 194–203.

[19]. Victoria Brown, E. D. (2018). Strengths and Weaknesses of Plagiarism Detection Software. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

[20]. Walpole, R. E. (1995). Pengantar Statistika. In *Pengantar Statistika Edisi Ke-3* (3rd ed., p. 514). PT Gramedia Pustaka Utama.

[21]. Wibowo, R. K., & Hastuti, K. (2016). Penerapan Algoritma Winnowing Untuk Mendeteksi Kemiripan Teks pada Tugas Akhir Mahasiswa. *Techno.Com*, *15*(4), 303–311.

[22]. Yusuf, B., Vivianie, S., Marsya, J. M., & Sofyan, Z. (2019). Analisis Perbandingan Algoritma Rabin-Karp dan Ratcliff / Obershelp untuk Menghitung Kesamaan Teks dalam Bahasa Indonesia. *Seminar Nasional APTIKOM (SEMNASTIK)*, 61–69.

[23]. Analisis,*"Analisis"*, https://kbbi.web.id/analisis, accessed on 01 January 2020, at 09:19 WIB

[24]. Metode Penelitian Kualitatif, *"*Metode Penelitian Kualitatif" https://www.gurupendidikan.co.id/metode-penelitian-kualitatif/, accessed on 25 January 2020, at 14:52 WIB