**Indonesian Journal of Electrical Engineering and Computer Science**
Vol. 29, No. 1, January 2023, pp. 572~582
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v29.i1.pp572-582 

# Automatic essay assessment in e-learning using winnowing algorithm

**Eka Larasati Amalia, Vivin Ayu Lestari, Vivi Nur Wijayaningrum, Ali Ar Ridla**
Department of Information Technology, Politeknik Negeri Malang, Malang, Indonesia

**Article Info**

***Article history:***

Received Apr 12, 2022
Revised Oct 7, 2022
Accepted Oct 11, 2022

***Keywords:***

Education
Jaccard coefficient
Online course
Similarity
Text mining

**ABSTRACT**

The pandemic has caused almost all educational institutions to use online learning media to support learning activities. E-learning is a technology that is widely used because it can accommodate all learning activities. However, in general, e-learning can only perform automatic assessments for multiple choice answers but not for essay answers, so that manual assessment by the teacher becomes difficult and takes a long time. In this study, the winnowing algorithm was applied to the automatic assessment process on students' essay answers by measuring their similarity to the teacher's answer key. The stages in the automatic assessment using the winnowing algorithm begin with forming a series of k-grams, calculating the hash value, forming a window from the hash value, calculating the fingerprint value, and calculating the Jaccard Coefficient to obtain the percentage of text similarity results. The test results show that the winnowing algorithm can provide good performance when the answers to questions are in the form of short entries with the number of hashes not smaller than the window value. Meanwhile, on questions with long answers, the winnowing algorithm can still work well with an average difference of 5.2% from the results of the assessment carried out by the teacher.



***Corresponding Author:***

Eka Larasati Amalia
Department of Information Technology, Politeknik Negeri Malang
Soekarno Hatta 9 Street, Malang 65145, Indonesia
Email: eka.larasati@polinema.ac.id

## 1. INTRODUCTION

The use of e-learning in the learning process is a form of adjustment to the development of information technology. Several educational institutions have developed e-learning systems with various features that can support the ease of the learning process since before the pandemic until now [1]. With e-learning, the distribution of material from teachers and students can be easily carried out and coordinated. In addition, students can also collect assignments and work on exam questions on the e-learning platform. In essence, e-learning can be used by students and teachers to carry out all learning activities integrated into a system [2]–[4], so that e-learning makes the activities of the learning process more efficient, creates a pleasant learning experience for students, and increases their learning activities [5]–[7].

The e-learning existence makes it easy for teachers to evaluate student learning outcomes through taking exams or assignments. Exam questions can be arranged in various forms, such as multiple-choice, true or false, short answers, and essays. Of course, the mechanism for assessing student learning outcomes adjusts to the type of questions given to students. Essay questions are considered to be the most helpful technique for measuring student learning outcomes. Essay questions are considered as a measuring tool for students' ability to memorize, remember, analyze, and write the results of their thoughts [8]. With essays, students can demonstrate their skills and knowledge related to the learning material that has been received [9]. However,

the assessment of essay answers is quite difficult for teachers to do because the process takes a lot of time, especially if the number of questions and students in a class is quite large, causing more questions to be assessed by the teacher. This condition can cause the quality of the assessment to decrease, and it is difficult to make an objective assessment [10].

The automatic assessment feature can be applied to e-learning to overcome these problems. Thus, the assessment of student answers in the form of essays can be done by the system. Several techniques have been applied by previous researchers related to this problem, such as Rabin Karp, latent semantic analysis (LSA), and Cosine Similarity. Rabin Karp is an algorithm that can be used to measure the level of similarity of text. This algorithm can be used to check the plagiarism level of a document [11]. Rabin Karp works by matching strings using a hash function as a comparison between search strings and substrings in the text. If the values of the two hashes are the same, then a further comparison of the characters will be carried out [12]. However, the weakness of Rabin Karp is its inability to perform a single pattern search process. The hash calculation on Rabin-Karp only calculates the number of hashes that have the same value in both documents [13]. Thus, if answers have the same meaning but the choice of words is not the same, it will produce a low level of similarity.

LSA as one of the techniques that can be used to carry out automatic assessments is a relatively simple method but has a reasonably high level of correlation [14]. Besides being used for automatic assessment [14], [15], LSA can also be applied to check plagiarism [16] and text summaries [17], [18]. Unfortunately, the LSA only assesses these documents by the terms in each document, so the word order or the layout of the terms is not taken into account. Even though this indirectly also affects the meaning contained in each document [19]. Another technique that is also often used to perform automated research on essay answers is Cosine Similarity. Cosine Similarity provides competitive results, as well as lower levels of complexity and faster load times compared to LSA [20]. As one of the popular methods to be applied to the calculation of document similarity, Cosine Similarity cannot be affected by the length or shortness of a document because the term value of each document is considered principal [21], [22]. Cosine Similarity has proven to be successful in conducting automatic essay assessments using *N*-Gram [23] and WordNet [24].

Another algorithm that is often used to solve problems related to text processing is winnowing. Assessment of essay answers can be done automatically by matching the strings of each word being compared. This algorithm can increase the efficiency of the document comparison process because of the concept of a window so that a hash value is formed, which is then used in the document matching process. In addition, the winnowing algorithm is also proven to provide a much greater accuracy value than Rabin Karp and Knuth Morris Pratt when used to detect plagiarism in documents [25].

The relationship between the parameters in the winnowing algorithm, such as *k*-grams, windows, and prime numbers, with the level of similarity, has not been clarified, so it is still unclear what effect the value will have on the results of the similarity measurement given. Therefore, the purpose of this study is to apply the winnowing algorithm with various combinations of parameter values to determine its impact on the results of automatic assessment of student answers in e-learning. By developing the automatic assessment feature in e-learning, objectivity and time efficiency in assessing essay answers are expected to be improved. Thus, the proposed e-learning system can accommodate various features that can make it easier for teachers and students to carry out the online learning process during the pandemic. In addition, it is easier for teachers to evaluate student learning outcomes in the form of assessment assignments, quizzes, or student exams that are carried out automatically in e-learning.

## 2. NOVELTY

The COVID-19 pandemic has had a major impact on the world of education. Learning from home using online media is a solution so that the teaching and learning process continues to run well. Adequate distance learning (e-learning) will provide great opportunities for students and educators to acquire knowledge and improve skills. E-learning can be defined as a computer-based online learning process used by students to review lecture materials, discussion forums, lecturer assessments, exam questions without reducing class face-to-face time [26]. E-learning can be presented in various forms of text, sound, video, images, and animation. E-learning can be accessed by students regardless of time and place [27], thus giving students more opportunities to repeat the existing material [28]. However, generally, e-learning developed by several previous researchers only had standard features consisting of discussion forums, uploading and downloading files, email notifications, student progress reviews, and search features [29].

This study offers e-learning in a single learning media portal to support and facilitate the learning process in Vocational High Schools. The e-learning features upload/download materials, online chat/discussion, automatic assessment of student evaluations (multiple choice and essay), real-time reporting of student grades, and attendance. Upload/download this material will be differentiated by subject and major, so students can only access those according to their major. Online chat/discussion is used to make it easier for students and teachers to discuss in real-time during learning hours according to the material at the meeting.

This automatic assessment of student evaluation is used to make it easier for teachers to evaluate multiple choice answers or essays automatically. In addition to the multiple-choice automatic assessment model, in this case, the researcher also focuses on essay questions. The automatic essay answer scoring system is a way that can facilitate and speed up the process of assessing essay answers so that teachers do not need to make corrections manually. Thus, it becomes easier for teachers to evaluate, specifically when assessing student practicum reports.

## 3. METHOD

In this study, the automatic assessment process for essay answers was carried out by implementing the winnowing algorithm. Winnowing algorithm is a method of finding similar words in documents by comparing fingerprints on documents [30]. The implementation of the winnowing algorithm requires input in the form of an answer text, then it will produce an output in the form of a percentage of the results of the similarity of the text. Figure 1 is a flowchart of the automatic assessment process for essay answers using the winnowing algorithm.

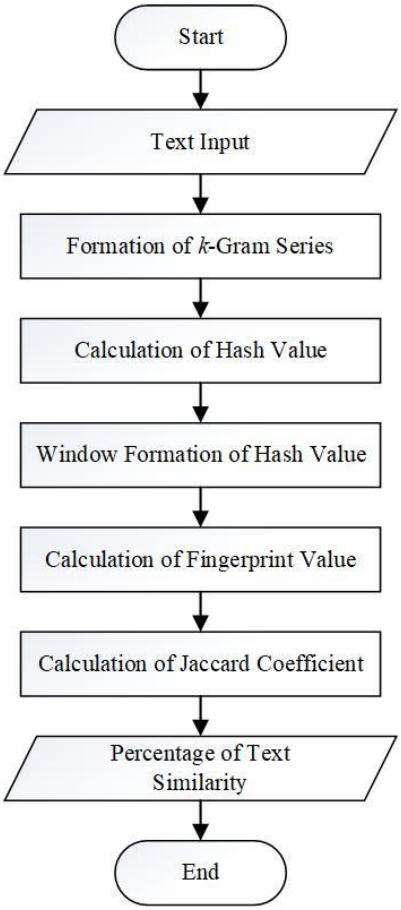


Figure 1. Flowchart of the proposed method

Based on Figure 1, text input is a character string that will be used as input to be processed in the winnowing algorithm. This input will be matched with the answer key text document to obtain a level of similarity, which will generate the correct answer value for the answer key automatically. The input of this algorithm is a text document that is processed to produce output in the form of a collection of hash values called fingerprints [31]. This fingerprint is used as a basis for comparison between the entered text file and the teacher's answer key so that the level of similarity of the text is obtained.

### 3.1. Formation of *k*-Gram series

In the winnowing algorithm, there is a *k*-gram formation which is a method for separating text by forming a substring along with *k* characters from a string [32]. The *k*-Gram is a series of terms with length *k*. The formation of the *k*-Gram series in the winnowing algorithm is done by forming a series of characters along with *k* from the results of removing irrelevant characters. A good *k* value is neither too small nor too large. The first *k*-Gram series starts from the 1st character to the *k*-th character, the second series starts from the 2nd character to the *k*+1 character, and so on until a *k*-Gram series of all characters is formed. Figure 2 shows an example of forming a *k*-Gram series in the text "*bunga merah*" (without quotes) with *k*=3.

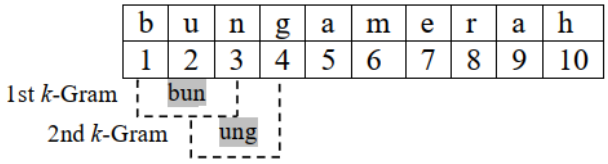


Figure 2. Formation of *k*-gram on "*bunga merah*"

With *k*=3 in the text "*bunga merah*", eight *k*-Grams are formed as:

- 1st *k*-Gram : bun
- 2nd *k*-Gram : ung
- 3rd *k*-Gram : nga
- 4th *k*-Gram : gam
- 5th *k*-Gram : ame
- 6th *k*-Gram : mer
- 7th *k*-Gram : era
- 8th *k*-Gram : rah

the comparison text used is "*bunga biru*" (without the quotes) so that the formation of the *k*-Gram series in the comparison text is shown in Figure 3.

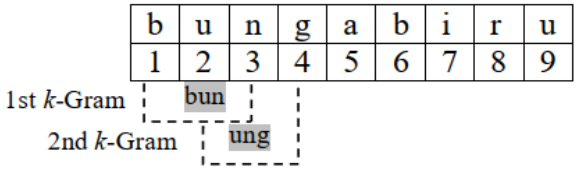


Figure 3. Formation of *k*-gram on "*bunga biru*"

With *k* = 3 in the text "*bunga biru*", seven *k*-Grams are formed as:

- 1st *k*-Gram : bun
- 2nd *k*-Gram : ung
- 3rd *k*-Gram : nga
- 4th *k*-Gram : gab
- 5th *k*-Gram : abi
- 6th *k*-Gram : bir
- 7th *k*-Gram : iru

### 3.2. Calculation of hash value

The hash function is to take string input and convert it to a fixed-length output string [33]. The winnowing algorithm uses a rolling hash to calculate the hash value of each gram series. The hash function with rolling hash is defined in (1), where *H* is the hash value, *c* is the character in grams, *b* is the base number, and *k* is the number of gram characters:

$$H_{c1\ldots ck} = c_1 \times b^{k-1} + c_2 \times b^{k-2} + \cdots + c_{k-1} \times b^k + c^k \quad (1)$$

for the second to the last hash value, the calculation does not need to be done using iterations one by one, but the hash value calculation can be done using (2).

$$H_{c2\ldots ck+1} = (H_{c1\ldots ck} - c_1 \times b^{k-1}) \times b + c_{k+1} \quad (2)$$

By using the results of the formation of the previous *k*-Gram series in Text 1 and Text 2 (comparison text), *b*=2, and *k*=5, the calculation of the hash value is shown in Table 1.

Table 1. The result of hash value calculation

| Text | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 | Value 6 | Value 7 | Value 8 |
|---|---|---|---|---|---|---|---|---|
| Text 1 | 736 | 791 | 895 | 1115 | 1547 | 2424 | 4161 | 7642 |
| Text 2 | 736 | 791 | 895 | 1104 | 1529 | 2388 | 4109 | |

### 3.3. Calculation of hash value

The winnowing algorithm does not use all the hash values of each gram series that has been formed. The hash value formed in the previous step will be divided into a window of size *w*. The first window contains the first hash values up to *w*. The second window is formed from the second hash value to *w*+1 and so on until a window is formed from all hash values. The window formation from the hash value calculation results in the previous stage with the window width (*w*)=4 is shown in Figure 4 for the Text 1 and Figure 5 for the Text 2. The window that is formed in the Text 1 is shown in Table 2. Meanwhile, Table 3 is the result of window formation in the Text 2.

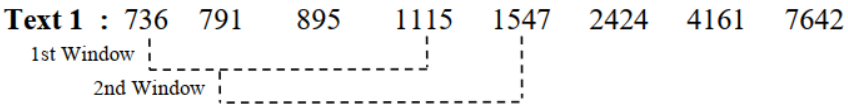


Figure 4. Window formation on text 1

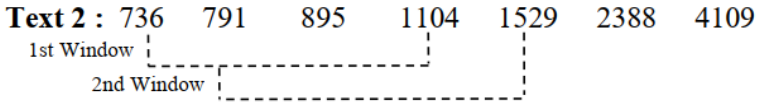


Figure 5. Window formation on text 2

Table 2. The result of window formation on text 1

| W1 | W2 | W3 | W4 |
|---|---|---|---|
| 736 | 791 | 895 | 1115 |
| 791 | 895 | 1115 | 1547 |
| 895 | 1115 | 1547 | 2424 |
| 1115 | 1547 | 2424 | 4161 |
| 1547 | 2424 | 4161 | 7642 |

Table 3. The result of window formation on text 2

| W1 | W2 | W3 | W4 |
|---|---|---|---|
| 736 | 791 | 895 | 1104 |
| 791 | 895 | 1104 | 1529 |
| 895 | 1104 | 1529 | 2388 |
| 1104 | 1529 | 2388 | 4109 |

### 3.4. Calculation of fingerprint value

The fingerprint is a technique that aims to prevent unauthorized copying of digital content, thus making digital content less likely to be duplicated [30]. The fingerprint value is determined by selecting the smallest hash value from each window. The selection of the fingerprint value from the window formation results in the previous stage is shown in Table 4 and Table 5.

Table 4. The result of fingerprint value calculation on text 1

| W1 | W2 | W3 | W4 |
|---|---|---|---|
| 736 | 791 | 895 | 1115 |
| 791 | 895 | 1115 | 1547 |
| 895 | 1115 | 1547 | 2424 |
| 1115 | 1547 | 2424 | 4161 |
| 1547 | 2424 | 4161 | 7642 |

Table 5. The result of fingerprint value calculation on text 2

| W1 | W2 | W3 | W4 |
|---|---|---|---|
| 736 | 791 | 895 | 1104 |
| 791 | 895 | 1104 | 1529 |
| 895 | 1104 | 1529 | 2388 |
| 1104 | 1529 | 2388 | 4109 |

### 3.5. Calculation of Jaccard coefficient

This step is done by calculating the hash value and selecting the smallest fingerprint of the two text documents [34]. The fingerprint value formed from the winnowing algorithm is used to measure the percentage of similarity of the text using the Jaccard Coefficient, as shown in (3).

$$Similarity(d_i, d_j) = \frac{|w(d_i) \cap w(d_j)|}{|w(d_i) \cup w(d_j)|} \times 100\% \quad (3)$$

In (3), $d_i$ and $d_j$ are the fingerprint values in the text, $w(d_i) \cap w(d_j)$ is the number of fingerprint values that are the same between the *i*-th text and *j*-th text, while $w(d_i) \cup w(d_j)$ is the total fingerprint value of the *i*-th text and the *j*-th text. The results of the similarity calculation using the Jaccard Coefficient equation for Text 1 and Text 2 are as:

Text 1: [736][791][895][1115][1547]
Text 2: [736][791][895][1104]

$$\begin{aligned} Similarity(d_i, d_j) &= \frac{[739][791][895]}{[739][791][895][1115][1547][1104]} \times 100\% \\ &= \frac{3}{6} \times 100\% \\ &= 50\% \end{aligned}$$

based on the results of these calculations, it is known that the percentage of text similarity between Text 1 and Text 2 is 50%.

## 4. RESULTS AND DISCUSSION

In this study, two testing mechanisms were carried out, which consisted of testing the combination of *k*-gram values, windows, and prime numbers and testing the accuracy of the system assessment results based on the teacher's assessment. The first test aims to determine the combination of *k*-gram values, windows, and prime numbers to obtain the right combination to produce good results. The second test aims to measure the accuracy of the results of the assessment system with the assessment carried out by the teacher manually.

### 4.1. Testing the combination of *k*-gram values, window, and prime numbers

The testing mechanism is carried out by comparing the similarity of the two answers. The question and answers used in this test are:

- Question: "Explain the meaning of metamorphosis?"
- First answer : "Metamorphosis is a biological developmental process in animals that involves changes in physical appearance and structure after birth or hatching. These physical changes occur as a result of cell growth and cell differentiation which are radically different."
- Second answer: "Metamorphosis is the process of a biological development of animals with changes in appearance and structure after birth or hatching. These physical changes are caused by cell proliferation and differentiation of very different cells."

By using ten combinations of *k*-gram values, windows, and prime numbers, the test results are shown in Table 6 and Figure 6. The combination of parameter values in this test was adopted from several previous studies [35]–[37], which resulted in a fairly high accuracy for similar cases. These values are used in this study

to measure their performance in the case of automatic assessment of student answers, with some adjustment of parameter values.

Table 6. The results of the combination of parameters on one question

| Experiment No. | *k*-gram | Window | Prime numbers | Similarity percentage |
|---|---|---|---|---|
| 1 | 3 | 4 | 7 | 56.54% |
| 2 | 3 | 3 | 7 | 57.31% |
| 3 | 4 | 3 | 7 | 49.63% |
| 4 | 4 | 7 | 3 | 44.57% |
| 5 | 6 | 7 | 3 | 35.74% |
| 6 | 2 | 4 | 7 | 74.04% |
| 7 | 2 | 7 | 7 | 75.22% |
| 8 | 2 | 7 | 19 | 72.96% |
| 9 | 4 | 5 | 5 | 47.08% |
| 10 | 4 | 7 | 19 | 48.33% |

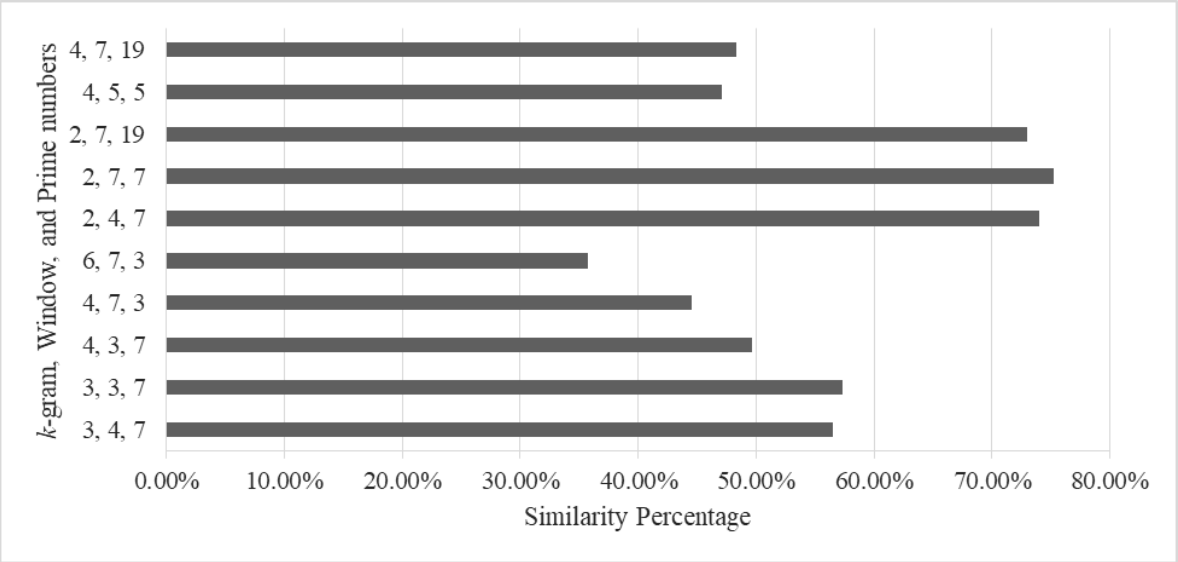


Figure 6. Graph of parameter combination test results

Separately, the test was carried out again on ten essay questions with various combinations of parameters shown in Table 7. While the test results are shown in Figure 7. Based on the test results shown in Figure 6 and Figure 7 using various combinations of parameters, it is known that the smaller the *k*-gram value, the greater the percentage value, which indicates the high similarity of the text. A small *k*-gram value indicates the more detailed the checking is carried out on each word. The *k*-gram value functions as a word splitter into a number according to the *k*-gram value. However, too small a *k*-gram value causes the duration of checking each word or sentence to be longer because more and more calculations are carried out so that it can increase the computation time.

The window value also affects the percentage of text similarity. The longer the text and the smaller the *k*-gram, the greater the percentage of similarity of the text. Meanwhile, the larger the prime number, the less different the result.

Table 7. Combination of *k*-gram values, window, and prime numbers

| Parameter | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 | Experiment 5 |
|---|---|---|---|---|---|
| *k*-gram | 3 | 3 | 4 | 4 | 3 |
| window | 4 | 4 | 4 | 3 | 2 |
| prime numbers | 2 | 5 | 5 | 5 | 3 |

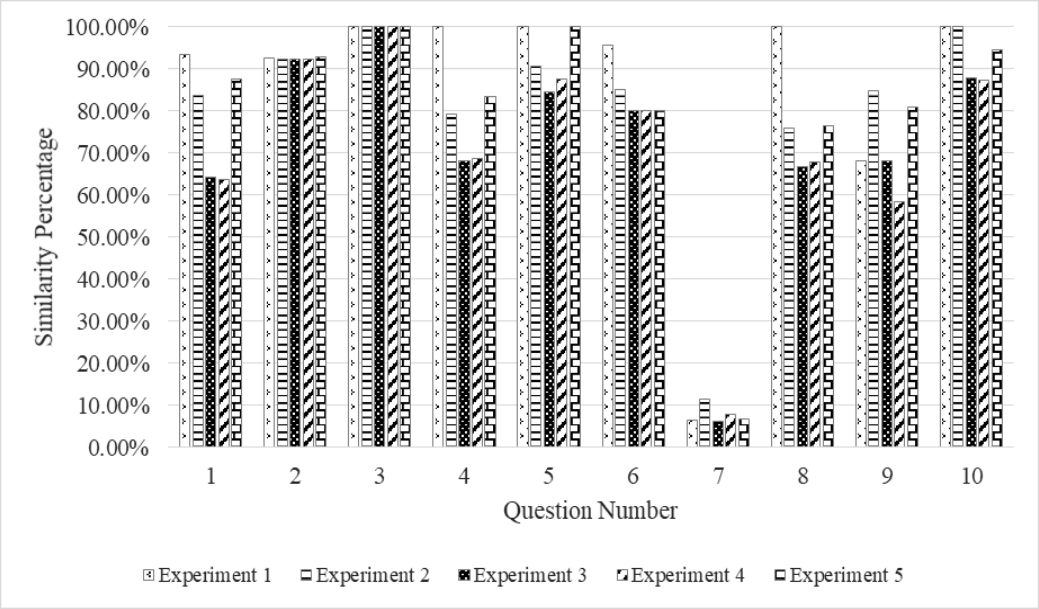


Figure 7. The results of testing the combination of *k*-gram values, window, and prime numbers

### 4.2. Accuracy testing based on teacher's assessment

The evaluation is carried out by comparing student answers with the teacher's answer key. This test uses 20 answers, consisting of ten student answers from working on Indonesian language questions and ten teacher answer keys. The test results are shown in Table 8.

Table 8. The result of accuracy test

| No | Student ID | Assessment by system | Assessment by teacher |
|---|---|---|---|
| 1 | 201#3 | 63 | 56 |
| 2 | 202#2 | 92 | 90 |
| 3 | 202#1 | 100 | 90 |
| 4 | 201#8 | 68 | 65 |
| 5 | 201#4 | 87 | 82 |
| 6 | 202#3 | 80 | 78 |
| 7 | 202#5 | 67 | 67 |
| 8 | 200#3 | 58 | 50 |
| 9 | 200#6 | 87 | 79 |
| 10 | 202#4 | 87 | 80 |
| Total | | 789 | 737 |

Based on the results of the accuracy-test in Table 8, it is known that the difference between the assessment system and the teacher's assessment is 52. Thus, it can be concluded that the level of accuracy produced is quite good, with an average difference of 5.2%. Winnowing algorithm is good enough to be implemented in short essay questions and still stable for long essay answers.

### 4.3. Comparison of winnowing with other algorithms

Through the test results, using the value of *k*-gram=3, the winnowing algorithm can provide a similarity level of 83.67%. The results showed that the student's answer correction system was running well and gave the desired results. When compared with the Robin Karp algorithm implemented in previous studies for string matching, the similarity level is 71.13% when using *k*-gram=3 [38]. The Cosine Similarity algorithm has also been used to check document similarity by producing a similarity level of 23% [39]. While the Knuth Morris Pratt algorithm provides an accuracy value of 9.20% to detect the similarity of two documents [25]. The comparison of the accuracy of the winnowing algorithm and several other algorithms used to check the similarity of the two documents can be seen in Table 9. The results of this test prove that the winnowing algorithm is able to detect the level of text similarity with a higher level of accuracy than the other three algorithms. This is because the winnowing algorithm is quite stable and effective as string matching.

Table 9. Comparison of algorithm accuracy

| Algorithm | Accuracy |
|---|---|
| Winnowing | 83.67% |
| Rabin Karp | 71.13% |
| Cosine Similarity | 23.00% |
| Knuth Morris Pratt | 9.20% |

## 5. CONCLUSION

The e-learning system was successfully developed using the Laravel framework with several features such as logging in, uploading and downloading learning materials, uploading and downloading assignments, creating and filling out attendance forms, and creating and working on exam questions. The existence of an automatic assessment feature in e-learning is proven to make it easier for teachers to evaluate students' essay questions so that the assessment process can be carried out effectively and efficiently. The winnowing algorithm for the assessment process of essay questions can give good performance on short essay answers with a hash number not smaller than the window value.

When used for long essay answers, the winnowing algorithm can still work well with an average difference of 5.2% from the actual assessment results. The *k*-gram value is proven to affect the similarity results, where the smaller the k-gram value, the greater the value obtained by students. On the other hand, the greater the value of *k*-gram, the smaller the similarity value, which causes the smaller the value obtained by students. In addition, the window value can also affect the same number of fingerprints, where the greater the window value, the more hash values are accommodated, so that the more the number of fingerprints are the same and can affect the similarity value. The winnowing algorithm is proven to provide the highest accuracy value compared to Rabin Karp, Cosine Similarity, and Knuth Morris Pratt when used to detect the level of document similarity, with an accuracy value of 83.7%.

## REFERENCES


[1] I. W. K. Wati, A. S. Sari, Widodo, and R. Setyaningsih, "Media need analysis of learning practicum in the COVID-19 pandemic," *Journal of Mechanical Engineering Education*, vol. 5, no. 2, pp. 155–162, 2020, doi: 10.30870/vanos.v5i2.9441.

[2] L. T. Stickney, R. F. Bento, A. Aggarwal, and V. Adlakha, "Online higher education: faculty satisfaction and its antecedents," *Journal of Management Education*, vol. 43, no. 5, pp. 509–542, 2019, doi: 10.1177/1052562919845022.

[3] M. S. Andrade, R. M. Miller, M. B. Kunz, and J. M. Ratliff, "Online learning in schools of business: the impact of quality assurance measures," *Journal of Education for Business*, vol. 95, no. 1, pp. 37–44, 2019, doi: 10.1080/08832323.2019.1596871.

[4] E. L. Amalia, A. P. Kirana, V. A. Lestari, V. N. Wijayaningrum, A. Shofiah, and R. U. Agustin, "Application of learning management system to improve teaching and learning activities in vocational high schools," in *Proceedings - IEIT 2021: 1st International Conference on Electrical and Information Technology*, 2021, pp. 63–68, doi: 10.1109/IEIT53149.2021.9587448.

[5] H. Aldowah, S. Ghazal, I. Naufal Umar, and B. Muniandy, "The impacts of demographic variables on technological and contextual challenges of e-learning implementation," in *Journal of Physics: Conference Series*, 2017, vol. 892, no. 1, doi: 10.1088/1742-6596/892/1/012013.

[6] N. Luo, Y. Zhang, and M. Zhang, "Retaining learners by establishing harmonious relationships in e-learning environment," *Interactive Learning Environments*, vol. 27, no. 1, pp. 118–131, 2019, doi: 10.1080/10494820.2018.1506811.

[7] S. Sukendro *et al.*, "Using an extended technology acceptance model to understand students' use of e-learning during COVID-19: Indonesian sport science education context," *Heliyon*, vol. 6, no. 11, 2020, doi: 10.1016/j.heliyon.2020.e05410.

[8] A. Alzahrani, A. Alzahrani, F. K. Al Arfaj, K. Almohammadi, and M. Alrashidi, "AutoScor: an automated system for essay questions scoring," *International Journal of Humanities Social Sciences and Education*, vol. 2, no. 5, pp. 182–187, 2015.

[9] K. Zupanc and Z. Bosnić, "Advances in the field of automated essay evaluation," *Informatica*, vol. 39, no. 4, pp. 383–395, 2015.

[10] L. Handayani, W. O. Alika, B. S. Negara, and Febiyanto, "A latent semantic analysis method for automatic scoring system at essay test," in *Journal of Physics: Conference Series*, 2020, vol. 1566, no. 1, doi: 10.1088/1742-6596/1566/1/012119.

[11] R. E. Putri and A. P. U. Siahaan, "Examination of Document similarity using Rabin-Karp algorithm," *International Journal of Recent Trends in Engineering and Research*, vol. 3, no. 8, pp. 196–201, 2017, doi: 10.23883/ijrter.2017.3404.4sndk.

[12] H. Herman, L. Syafie, T. Tasmil, and M. Resha, "Rabin-Carp implementation in measuring simalirity of research proposal of students," *Journal of Information Technology and Its Utilization*, vol. 3, no. 1, pp. 9-13, 2020, doi: 10.30818/jitu.3.1.3210.

[13] A. P. U. Siahaan *et al.*, "Combination of levenshtein distance and Rabin-Karp to improve the accuracy of document equivalence level," *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 2 special issue 27, pp. 17–21, 2018, doi: 10.14419/ijet.v7i2.27.12084.

[14] N. LaVoie, J. Parker, P. J. Legree, S. Ardison, and R. N. Kilcullen, "Using latent semantic analysis to score short answer constructed responses: automated scoring of the consequences test," *Educational and Psychological Measurement*, vol. 80, no. 2, pp. 399–414, 2020, doi: 10.1177/0013164419860575.

[15] R. S. Citawan, V. C. Mawardi, and B. Mulyawan, "Automatic essay scoring in e-learning system using LSA method with N-Gram fature for Bahasa Indonesia," in *The 3rd International Conference on Electrical Systems, Technology and Information (ICESTI 2017)*, 2018, vol. 164, p. 01037, doi: 10.1051/matecconf/201816401037.

[16] G. Cosma and M. Joy, "An approach to source-code plagiarism detection and investigation using latent semantic analysis," *IEEE Transactions on Computers*, vol. 61, no. 3, pp. 379–394, 2012, doi: 10.1109/TC.2011.223.

[17] H. Li, Z. Cai, and A. C. Graesser, "Computerized summary scoring: crowdsourcing-based latent semantic analysis," *Behavior Research Methods*, vol. 50, no. 5, pp. 2144–2161, 2018, doi: 10.3758/s13428-017-0982-7.

[18] R. Rani and D. K. Lobiyal, "A weighted word embedding based approach for extractive text summarization," *Expert Systems with Applications*, vol. 186, 2021, doi: 10.1016/j.eswa.2021.115867.

[19] M. M. Islam and A. S. M. L. Hoque, "Automated essay scoring using generalized latent semantic analysis," *Journal of Computers*, vol. 7, no. 3, pp. 616–626, 2012, doi: 10.4304/jcp.7.3.616-626.
[20] E. S. Pramukantoro and M. A. Fauzi, "Comparative analysis of string similarity and corpus-based similarity for automatic essay scoring system on e-learning gamification," in *2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016*, 2017, pp. 149–155, doi: 10.1109/ICACSIS.2016.7872785.
[21] S. Sohangir and D. Wang, "Improved SQRT-Cosine Similarity measurement," *Journal of Big Data*, vol. 4, no. 1, pp. 1–13, 2017, doi: 10.1186/s40537-017-0083-6.
[22] W. Darmalaksana, C. Slamet, W. B. Zulfikar, I. F. Fadillah, D. S. A. Maylawati, and H. Ali, "Latent semantic analysis and Cosine Similarity for hadith search engine," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 1, pp. 217–227, 2020, doi: 10.12928/TELKOMNIKA.V18I1.14874.
[23] M. A. Fauzi, D. C. Utomo, E. S. Pramukantoro, and B. D. Setiawan, "Automatic essay scoring system using N-GRAM and Cosine Similarity for gamification based elearning," in *ACM International Conference Proceeding Series*, 2017, pp. 151–155, doi: 10.1145/3133264.3133303.
[24] K. Jayakodi, M. Bandara, I. Perera, and D. Meedeniya, "WordNet and cosine similarity based classifier of exam questions using bloom's taxonomy," *International Journal of Emerging Technologies in Learning*, vol. 11, no. 4, pp. 142–149, 2016, doi: 10.3991/ijet.v11i04.5654.
[25] I. Widaningrum, D. Mustikasari, R. Arifin, and H. A. Pratiwi, "Evaluation of the accuracy of winnowing, Rabin Karp and Knuth Morris Pratt algorithms in plagiarism detection applications," in *Journal of Physics: Conference Series*, 2020, vol. 1517, no. 1, doi: 10.1088/1742-6596/1517/1/012093.
[26] A. Abdulazeez and S. Zeebaree, "Design and implementation of electronic learning system for Duhok Polytechnic University," *Academic Journal of Nawroz University*, vol. 7, no. 3, pp. 249–258, 2018, doi: 10.25007/ajnu.v7n3a250.
[27] S. Dhawan, "Online learning: a panacea in the time of COVID-19 crisis," *Journal of Educational Technology Systems*, vol. 49, no. 1, pp. 5–22, 2020, doi: 10.1177/0047239520934018.
[28] A. Alenezi, "The role of e-learning materials in enhancing teaching and learning behaviors," *International Journal of Information and Education Technology*, vol. 10, no. 1, pp. 48–56, 2020, doi: 10.18178/ijiet.2020.10.1.1338.
[29] A. Febliza and O. Okatariani, "The development of online learning media by using moodle for general chemistry subject," *Journal of Educational Science and Technology (EST)*, vol. 6, no. 1, pp. 40–47, 2020, doi: 10.26858/est.v6i1.12339.
[30] Y. Nurdiansyah and F. N. Muharrom, "Implementation of winnowing algorithm based K-gram to identify plagiarism on file text-based document," *Journal of Engineering and Applied Sciences*, vol. 13, no. special issue 10, pp. 8119–8125, 2018, doi: 10.3923/jeasci.2018.8119.8125.
[31] N. Ulinnuha, M. Thohir, D. C. R. Novitasari, A. H. Asyhar, and A. Z. Arifin, "Implementation of winnowing algorithm for document plagiarism detection," *Proceeding of the Electrical Engineering Computer Science and Informatics*, vol. 5, no. 5, pp. 631–636, 2018, doi: 10.11591/eecsi.v5i5.1599.
[32] A. P. . Siahaan, R. Rahim, M. Aan, and D. Siregar, "K-Gram as a determinant of plagiarism level in Rabin-Karp algorithm," *International Journal of Scientific & Technology Research*, vol. 6, no. 7, pp. 350–353, 2017, doi: 10.31219/osf.io/yxjnp.
[33] F. Zhao and Q. Liu, "A string matching algorithm based on efficient hash function," in *Proceedings - 2009 International Conference on Information Engineering and Computer Science, ICIECS 2009*, 2009, doi: 10.1109/ICIECS.2009.5363191.
[34] S. Schleimer, D. S. Wilkerson, and A. Aiken, "Winnowing: local algorithms for document fingerprinting," *SIGMOD '03: Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, 2003, pp. 76–85, doi: 10.1145/872757.872770.
[35] E. G. Hasan, A. Wicaksana, and S. Hansun, "The implementation of winnowing algorithm for plagiarism detection in moodle-based e-learning," in *Proceedings - 17th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2018*, 2018, pp. 321–325, doi: 10.1109/ICIS.2018.8466429.
[36] A. Rahmatulloh, N. I. Kurniati, I. Darmawan, A. Z. Asyikin, and J. D. Witarsyah, "Comparison between the stemmer porter effect and nazief-adriani on the performance of winnowing algorithms for measuring plagiarism," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 4, pp. 1124–1128, 2019, doi: 10.18517/ijaseit.9.4.8844.
[37] I. Fakhruddin and I. G. Anugrah, "Implementation of winnowing algorithm and simple additive weighting SAW for publication reference journal search system," *Journal of Development Research*, vol. 5, no. 2, pp. 61–72, Nov. 2021, doi: 10.28926/jdr.v5i2.141.
[38] A. P. U. Siahaan, "Rabin-Karp elaboration in comparing pattern based on hash data," *International Journal of Security and Its Applications*, vol. 12, no. 2, pp. 59–66, 2018, doi: 10.14257/ijsia.2018.12.2.06.
[39] D. Gunawan, C. A. Sembiring, and M. A. Budiman, "The implementation of cosine similarity to calculate text relevance between two documents," in *Journal of Physics: Conference Series*, 2018, vol. 978, no. 1, doi: 10.1088/1742-6596/978/1/012120.

## BIOGRAPHIES OF AUTHORS

**Eka Larasati Amalia** is a lecturer at the Department of Information Technology, Politeknik Negeri Malang. She graduated from Politeknik Elektronika Negeri Surabaya where she received her bachelor's degree. Then she received her master's degree in Electrical Engineering from Universitas Brawijaya. Her research interests focus on developing e-learning systems and virtual laboratories. She can be contacted at email: eka.larasati@polinema.ac.id.

**Vivin Ayu Lestari** obtained the bachelor's degree in Informatics Engineering from Universitas Negeri Malang and the master's degree of Computer Science from Universitas Brawijaya. She is currently a lecturer at the Department of Information Technology, Politeknik Negeri Malang. The focus of her research is on the field of information systems, especially the method of learning algorithms in the Java programming language. She can be contacted at email: vivin@polinema.ac.id.

**Vivi Nur Wijayaningrum** holds bachelor's and master's degrees from the Department of Computer Science, Universitas Brawijaya. Currently, she is a lecturer at the Department of Information Technology, Politeknik Negeri Malang. She has authored or co-authored several publications, with 8 H-indexes and over 200 citations. Her research interests include machine learning, decision support systems, and expert systems. She can be contacted at email: vivinurw@polinema.ac.id.

**Ali Ar Ridla** is an undergraduate student at the Department of Information Technology, Politeknik Negeri Malang. He is interested in website development and computer network. He can be contacted at email: aliridla11@gmail.com.