Comparison Analysis of Rabin-Karp and Winnowing Algorithms in Automated Essay Answer Assessment System

Titan Tawang Ilal Billhaqqi[a)], Galih Wasis Wicaksono[b)], Christian Sri Kusuma Aditya[c)]

Department of Informatics, Universitas Muhammadiyah Malang, Malang, Indonesia

Corresponding author: [a)]titan.tib@gmail.com
[b)]galih.w.w@umm.ac.id
[c)]christianskaditya@umm.ac.id

Abstract. The role of technology during the Covid-19 pandemic has been evidently dominant, particularly in the teaching and learning process. The application of learning from home, enables E-Learning to serve as the primary learning medium. E-learning requires an automatic essay answer assessment feature to improve the objectives and efficiency of online learning activities. This study proposes the automated assessment solutions utilizing Rabin-Karp and Winnowing algorithms. The types of essay questions handled in this study include the free description and limited description types. The answer key data and answer data will be further preprocessed to be similarly calculated to the two algorithms in the process. Algorithm accuracy assessment is conducted by converting the value into a human rates version compared with the evaluation conducted by lecturers manually. This study indicated that Rabin-Karp's Algorithm had better accuracy with the slightest difference of 27.81%, compared to Winnowing's Algorithm accuracy with a more significant distinction of 44.32%.

INTRODUCTION

The constantly evolving information technology is deemed beneficial for humans in overcoming learning during the current Covid-19 pandemic. Information technology plays an essential role in society to be able to keep doing things [1]. One of the benefits of utilizing information technology during the Covid-19 pandemic is in the online teaching and learning process without having to hold face-to-face interaction, thus reducing the potential spread of the Covid-19 virus.

One of the tools vitally required in online teaching and learning activities today is E-Learning. E-Learning becomes a tool, highly demanded by the teachers and students to support the presentation and delivery of knowledge and skills in the learning process [2]. E-Learning is predominantly utilized as a means of discussion between students and lecturers, in addition to a material presentation, assignment collection, questioning, and multiple-choice type answer assessment [3]. The correction or evaluation for essay-type answer assessment automatically in E-Learning is however inadequate, thereby continuing the manual assessment.

An essay or description test refers to a test that provides flexibility for students in expressing answers based on ideas, or students can organize their thoughts independently [4]. The essay question type is divided into the limited description and free description. A limited description type includes a form of question whose answer direction is limited so that the answers for this type are directed. The free description question type refers to a form of a question that allows students to freely answer their systematics or freely express opinions according to their ability [5].

Previous studies have examined the use of Rabin Karp and Winnowing algorithms. Optimization of Rabin-Karp with TF-IDF has shown an increased measurement of similarities between answers and key answers by an average of 11.81% [6]. The Winnowing Algorithm on automatic assessment of essay answers obtains the result that stable for long and short answer assessment. This study obtained the average accuracy of the difference between teacher and system assessments on each question of 5.683% [7]. Other studies have concluded that winnowing algorithms to measure similarities of thesis titles have a fast and accurate process [8]. Different studies have attempted to research the implementation ofrolling hashes in Rabin-Karp's Algorithm, which can improve answer detection results to be more accurate because rolling hashes can be changed based on characters discarded and added [9]. Research related to strengthening Winnowing algorithms with others, such as whitespace insensitivity in text detection algorithms, noise suppression, and position independence, has been conducted and can meet the necessary needs [10].

1st International Conference on Technology, Informatics, and Engineering
AIP Conf. Proc. 2453, 030018-1–030018-6; https://doi.org/10.1063/5.0095186
Published by AIP Publishing. 978-0-7354-4356-3/$30.00

This study will compare Winnowing and Rabin-Karp's algorithms to assess essay answers automatically. This research aims to fulfill of the automatic essay answer assessment feature in E-Learning, which is expected to improve objectivity and time efficiency in answering assessments. The study also attempts to improve the performance of Winnowing and Rabin Karp algorithms by utilizing hash values as parameters used in text matching. The data used in this study is sourced from E-Learning data in higher education, due to a wide variety and characteristics of essay answers.

METHODS

As described in Figure 1, stages in this study include data collection activities, preprocessing steps such as case folding, tokenizing, filtering, and stemming; Calculation of similarities with Rabin-Karp's Algorithm was compared to Winnowing's Algorithm; Convert values to human rates, and Testing results.

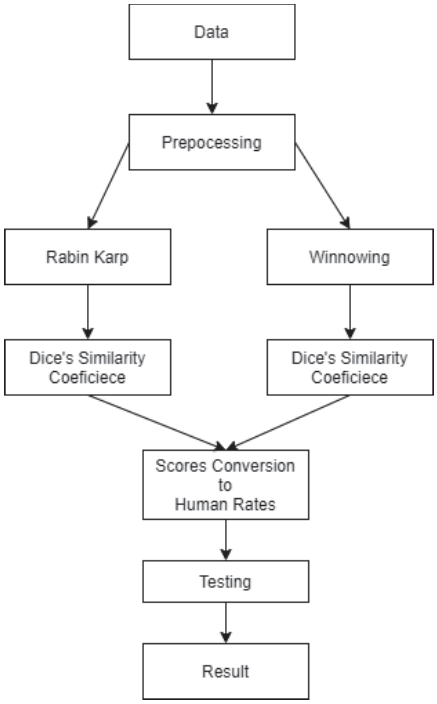

FIGURE 1. Research stage

The data collection procedure is conducted by logging in utilizing the lecturer's access rights at the E-Learning University of Muhammadiyah Malang (UMM) with the address of https://elmu.umm.ac.id. In this study, the courses utilized as data sources include Ethics and Professions. Questions and answers were undertaken from the quiz on communication materials. Three types of data must be collected: 1) essay question data, 2) key data essay answers, and 3) student answer data.

The preprocessing stage in this study refers to the main steps in text preprocessing, including case folding, tokenizing, filtering, and stemming [11][12]. In the filtering stage, the stop word list removal procedure is performed based on the specified stop word list. At this stage also conducted stemming by utilizing the library Sastrawi. Stemming refers to a stage to turn a word into a basic word by eliminating the affix from filtering results [13][14].

K-gram refers to a series of adjacent substrings based on the number of k, where the value at a k can be determined by the user, the formation of the substring is read from the beginning to the end of the document based on the value of k [15][16].

Hashing includes a unique value conversion in the formed string arrangement, obtained by calculating ASCII value in the string arrangement, also called fingerprint on a string [17]. The type of hashing used in Rabin-Karp and Winnowing is rolling hash [18], with the following Equation 1.

$$H_{(c_2 \ldots c_n)} = c_1 * b^{(n-1)} + c_2 * b^{(n-2)} + \cdots + c_{(n-1)} * b^{(n)} + c_n \quad (1)$$

In which:
c = ascii value of a character
b = prime number base (unspecified)
n = number of characters or length of n-gram circuit

Dice's similarity coefficients represent a stage aiming to find similarities between objects from variables covered between those objects [19][20]. Dice's similarity coefficients Equation 2 is presented as follows.

$$S = \frac{2C}{(A + B)} \tag{2}$$

In which:
S: Similarity
A: Number of K-Grams of Document 1
B: Number of K-Grams of document 2
C: Number of K-Grams with the same hash value in both documents

Rabin-Karp refers to an algorithm based on string matching, with the basic principle of finding pattern equations between text input and text comparison. Pattern similarity measurement is conducted by utilizing substring on a text utilizing hashing. If obtaining a lot of similarity of hash value on two documents will result in a high similarity value [21][22][23]. The stages in the Rabin-Karp Algorithm comprise from forming a series of grams with size k (K-Gram), then carried out the formation of hash values with the rolling hash function of each gram formed, and determine the equation of 2 documents with the equation dice's similarity coefficients [24].

The Winnowing Algorithm utilizes string matching to generate a fingerprint marker of a document based on the hash value formed. The previously developed hash value will be divided into a predetermined w-sized window [25][26][27]. Stages in Winnowing are conducted through the formation of a series of grams with a size k (K-Gram), then continued with the construction of hash values utilizing the rolling hash function of each gram formed. The last step is performed by dividing all hashes into a window and determining the equation of 2 documents with the equation dice's similarity coefficients [28][29].

The similarity result of dice's similarity coefficients will be converted based on the vulnerable human rate version value [30]. The scoring range is illustrated in Table 1.

TABLE 1. Human rate range

Value Range Comparison	
Similarity Value	Human Rates Value
0,01 – 0,10	10
0,11 – 0,20	20
0,21 – 0,30	30
0,31 – 0,40	40
0,41 – 0,50	50
0,51 – 0,60	60
0,61 – 0,70	70
0,71 – 0,80	80
0,81 – 0,90	90
0,91 – 1	100

At the testing stage, it is significantly important to measure the difference between the assessment by the system and the assessment manually [7]. Successful indicators are denoted from the small value of the difference between system assessment and manual assessment.

RESULTS AND DISCUSSION

The process of obtaining and collecting three types of essay data, answer keys, and student answers are conducted manually by copying each data from E-Learning. The essay question data is obtained from the quiz feature on the lecturer's account. The total number of questions in the communication material is eight questions. The key data of

essay answers are compiled from lecturer teaching materials, validated by the lecturers. The essay answer data was obtained from the answer field of 41 students, with a total of 328 answer items. The classification of questions has two types, that is free description types and limited description types.

The key data of the answer and the student's essay answer data will be conducted for text preprocessing. Furthermore, in which the two data are also calculated for their similarities utilizing Rabin-Karp and Winnowing methods. The calculation results of Rabin-Karp and Winnowing algorithms will be converted into human rates, that this step aimed to determine the final value of each answer.

In Figure 2, the average comparison of assessment results with Rabin-Karp and Winnowing algorithms is visualized with the results of lecturers' assessments. It appears that the results of the assessment system with the Rabin-Karp Algorithm have results closer to the evaluation of lecturers compared to Winnowing algorithms.

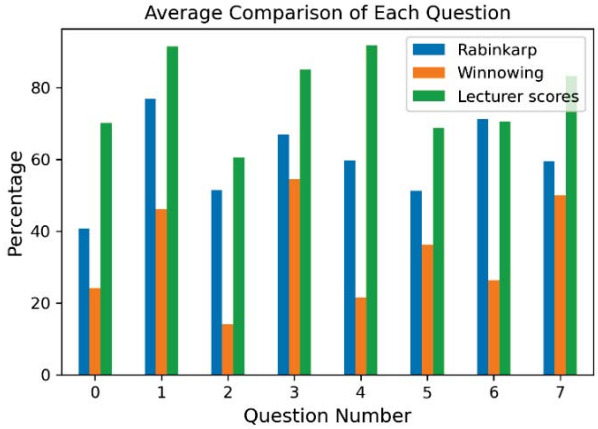

FIGURE 2. Average comparison of Rabin-Karp and Winnowing algorithms with lecturer grades

Analysis of each student's answer to each question item is based on the question types of free description types and limited description types. The rating with Winnowing's Algorithm mostly has a higher difference compared to Rabin-Karp's Algorithm, which occurs in the type of free description question, presenting a considerable difference in student essay answers to the answer key. The result is different from limited description types, where the appraisal results with two algorithms have a fairly identical difference result. The average comparison of assessment results with Rabin-Karp and Winnowing algorithms is visualized in Figure 3.

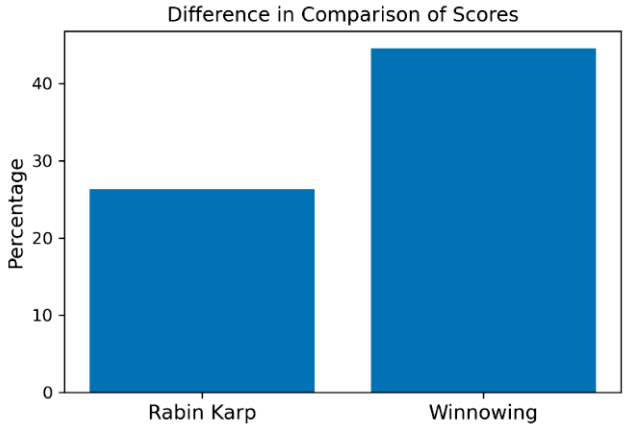

FIGURE 3. Comparison of Rabin-Karp and Winnowing scores

In general, the difference between lecturer assessment and system assessment utilizing the Rabin-Karp Algorithm is more negligible with a value of 27.81%, compared to the difference between lecturer assessment and Winnowing algorithm assessment result of 44.32%. Figure 3 indicates that Rabin-Karp algorithm accuracy has better performance than the Winnowing Algorithm.

CONCLUSIONS

Based on the steps and results of the conducted research, Rabin-Karp and Winnowing algorithms are proper to conduct automatic answer assessments. Both algorithms also indicate the same scoring results when used on limited description question types with specific answer types. However, both algorithms present a considerable difference in the scoring results regarding the free description question. Therefore, further study is encouraged to include the use of different words in the answer or use word synonyms to detect different words and phrases, presenting similar meaning in the answer to reduce the difference in judgment on the type of free description question and the type of limited description.

REFERENCES

1. R. Komalasari, "Manfaat Teknologi Informasi Dan Komunikasi Di Masa Pandemi Covid 19," Tematik, **7**, (1), 38–50 (2020).
2. A.-M. Tîrziu and C. Vrabie, "Education 2.0: E-Learning Methods," Procedia - Soc. Behav. Sci., **186**, 376–380 (2015).
3. W. Hartanto, "Penggunaan E-Learning sebagai Media Pembelajaran," J. Pendidik. Ekon., **10**, (1), 1–18 (2016).
4. N. K. R. Purwatia and I. W. Sumandya, "Penerapan Partial Credit Model (PCM) dalam Mengevaluasi Tes Uraian," Emasains, **VIII**, (1), 77–85 (2019).
5. A. H. Z. Wulan and R. Aristia, "Jenis - Jenis Instrumen dalam Evaluasi Pembelajaran," Univ. Muhammadiyah Sidoarjo, 1–13 (2018).
6. I. Print, S. Bahri, and R. Wajhillah, "Optimalisasi Algoritma Rabin Karp menggunakan TF-IDF Dalam Pencocokan Text Pada Penilaian Ujian Essay Otomatis," **4** (2020).
7. F. E. Kurniawati and W. M. Pradnya, "Implementasi Algoritma Winnowing Pada Sistem Penilaian Otomatis Jawaban Esai Pada Ujian Online Berbasis Web," **VI**, 169–175 (2020).
8. Nurdin and M. Amin, "Sistem Pendeteksian Kemiripan Judul Skripsi Menggunakan," J. Nas. Inform. dan Teknol. Jar., **2**, (1), 90–97 (2017).
9. J. Priambodo, "Pendeteksian Plagiarisme Menggunakan Algoritma Rabin-Karp," J. Inform. Univ. Pamulang, **3**, (1), 39–45 (2018).
10. N. ALAMSYAH, "Perbandingan Algoritma Winnowing Dengan Algoritma Rabin Karp Untuk Mendeteksi Plagiarisme Pada Kemiripan Teks Judul Skripsi," Technol. J. Ilm., **8**, (3), 124 (2017).
11. M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Utilizing Sastrawi," IOP Conf. Ser. Mater. Sci. Eng., **874**, (1) 2020.
12. T. Winarti, J. Kerami, and S. Arief, "Determining Term on Text Document Clustering utilizing Algorithm of Enhanced Confix Stripping Stemming," Int. J. Comput. Appl., **157**, (9), 8–13 (2017).
13. A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," Proc. 2016 4th Int. Conf. Cyber IT Serv. Manag. CITSM (2016).
14. A. Yudhana, S. -, and A. Djalil, "Implementation of Pattern Matching Algorithm for Portable Document Format," Int. J. Adv. Comput. Sci. Appl., **8**, (11), 509–512 (2017).
15. R. Sutoyo, I. Ramadhani, and A. D. Ardiatma, "Detecting Documents Plagiarism utilizing Winnowing Algorithm and K-Gram Method," 67–72.
16. D. B. Rahmawati, M. L. Irfani, and R. B. R. Purba, "Text Mining to Detect Plagiarism in E-Learning System Utilizing Rabin Karp Algorithm," IRE Journals, **3**, (8), 183–191 (2020).
17. A. Agung Putri Ratna, D. Lalita Luhurkinanti, I. Ibrahim, D. Husna, and P. Dewi Purnamasari, "Automatic Essay Grading System for Japanese Language Examination Utilizing Winnowing Algorithm," Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018, 565–569 (2018).
18. H. Jiang and S. J. Lin, "A Rolling Hash Algorithm and the Implementation to LZ4 Data Compression," IEEE Access, **8**, 35529–35534 (2020).
19. A. Sunyoto and T. Informatika, "Implementasi Algoritma Rabin Karp untuk Pendeteksian Plagiat Dokumen Teks Menggunakan Konsep Similarity," 23–28 (2013).
20. R. Kaderka et al., "Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients," Radiother. Oncol., **131**, 215–220 (2019).
21. A. P. U. Siahaan et al., "Combination of levenshtein distance and rabin-karp to improve the accuracy of document equivalence level," Int. J. Eng. Technol., **7**, (2 Special Issue 27), 17–21 (2018).