P-ISSN: 2776-7329, E-ISSN: 2776-7310
Proceedings
5th Vocational Education International Conference
Semarang, 13 July 2023

Automatic Short Answer Scoring (ASAS) Using String-based Similarity and Query Expansion

Feddy Setio Pribadi*, Uswatun Hasanah, Devi Nur Sa'diah

Department of Electrical Engineering, Universitas Negeri Semarang, Indonesi
Email: feddy.setio.p@mail.unnes.ac.id

ABSTRACT
Advances in technology and information in the field of education have made the learning system computerized, where the learning process and assessment process can be carried out using electronic media connected to the internet. The assessment process in e-learning is used to measure students' ability to understand the learning material obtained during the learning process. In e-learning, the assessment process for essay exams is still done manually by educators, which requires a lot of time and energy, so there is a need for an automatic essay exam assessment system. This research aims to assess the short answer essay exam using the string-based similarity method combined with the query expansion method with stages carried out including case folding, tokenizing, stop word, stemming and calculating similarity scores using the cosine and Jaccard methods and their modifications as well as understanding the meaning. Synonyms of the words being compared using the query expansion method. The results of the correlation test showed that the weighted cosine coefficient and weighted Jaccard coefficient methods combined with query expansion have the highest correlation value of 0.913 to 0.943, however, the comparison test shows that the weighted cosine coefficient method combined with query expansion produces a decision that there is no difference in average value - The average between the results of the educator's assessment and the results of the automatic short answer scoring program assessment using the weighted cosine coefficient method and the query expansion method.

Keywords: *Automatic short Answer Scoring, String Similarity and Query Expansion.*

1. INTRODUCTION

Assessment of student learning outcomes is used to measure students' ability to understand the learning material that has been obtained during the learning process [1]. Assessment of learning outcomes can take the form of subjective exams in the form of essay exams and objective exams in the form of multiple-choice exams [2].

The use of technology and information in the field of education makes the learning process and assessment process computerized, such as with e-learning [3][4]. In some e-learning, multiple-choice exams are used more often [5] because answer choices are provided [6] and the assessment process in multiple-choice exams is easier [7] but multiple-choice exams are less able to measure the level of students' ability in the knowledge they have acquired. [8]. Meanwhile, essay exams can measure students' abilities regarding the knowledge they have acquired [9] [10] because in essay exams there are no answer choices [11], so it requires students to answer by composing their own sentences [12]. Essay exams are rarely used in some e-learning because students' answers to essay exams can vary [13][14] so the assessment process on essay exams is still corrected manually by educators which requires a lot of time and energy [15][16]. Assessments carried out manually can be subjective [17] because the grades given by one educator and another educator who corrects them are sometimes different for each student on the same question [18], so that the assessment results are inconsistent and the quality of the assessment decreases [11], because there is a need for an essay exam assessment system that is carried out automatically [19][20][21]. Automatic essay scoring is a computer program created for the purpose of simplifying and speeding up the assessment process on essay exams [22] [23] which consists of two types, namely automatic essay scoring (AES) for assessing long answer essay exams and automatic short answer scoring (ASAS).) for marking on short answer essay exams [1][24][25]. This research focuses more on the automatic assessment of short answer essays to measure the similarity between the answer key and the student's answers being compared [26][27].

Previous research developed assessment of short essay exams using the string-based similarity method [1]

[18][24][28] because this method is more relevant for assessing short answer essays which have a limited number of words, but the string-based similarity method will only provide scores for the same words between the answer key and the student's answers are compared. In previous research [1], we have modified the string-based similarity method, namely cosine coefficient, Jaccard coefficient and dice coefficient into the weighted cosine coefficient, weighted Jaccard coefficient and weighted dice coefficient method by adding weighting to word terms which aims to increase the similarity score. Another study [28] also modified the blue method to M-blue to improve the similarity score results obtained by adding weighting to word terms based on the level of word importance.

The cosine coefficient, weighted cosine coefficient and Jaccard coefficient and weighted Jaccard coefficient methods used in the automatic short answer essay grading system do not pay attention to word order and only give scores for the same words [1][24][37] and cannot understand the words which are synonymous so that there is a need for other methods to deal with this problem, apart from that the use of one answer key also affects the score obtained, because only one answer key is compared with various student answers [24] so the score obtained is less than optimal. Therefore, this research uses more than one reference answer which is used as an answer key to handle varied student answers. This research also proposes a method to handle the problem of the semantic meaning of the words being compared by adding a query expansion method. The query expansion method is used to expand the query by adding additional query words to the information retrieval process [29] [30] and modifying the query to improve the query by changing words that are synonymous with synonyms [31] so that information needs are met and get better results [32]. The use of the query expansion method in this research utilizes the Indonesian thesaurus dictionary to enrich queries to search for and find synonyms of the words being compared [33]. So, the automatic assessment of short answer essays in this study uses the string-based similarity method combined with the query expansion method.

2. MATERIAL AND METHODS

2.1 Training Program in SMK

This research uses a data collection in the form of daily test results of short essays in Indonesian that do not have mathematical equations. The data used comes from Boedi Oetomo Cilacap Vocational School which consists of 5 short essay exam questions about knowledge of simulation and digital communication subjects, 3 reference answers used as answer keys, answer results from 2 classes, namely X Accounting and Financial Institutions (AKL) and X Automation and Office Governance (OTKP) with a total of 68 students and grades for each question along with the test scores conducted manually.

Table 1. Sample Data Set

Question	What is the function of rectangular symbols in a flow chart?
Reference 1	Used in processing data and calculation processes
Reference 2	To show every data processing that will be carried out by the computer
Student Answer	To indicate the occurrence of a calculation operation in an algorithm

2.1 Text Processing

In this research, we carried out a text preprocessing process in preparing the text to remove noise before carrying out the similarity score calculation process [34] which aims to change the input data in the form of reference answers and student answers into data that is ready to be processed at the next stage. The text preprocessing process consists of a case folding process which aims to change all the letters to lower case [35] then a tokenizing process is carried out to cut sentences into groups of words [10] followed by a stop word removal process to remove or delete unimportant words included in the stop list from the collection existing words [36] so that the results obtained are only important words which are then carried out a stemming process which aims to change the words into their basic word forms by removing the word affixes [37] by using a literary library for the stemming process on texts in Indonesian. The results of the text preprocessing process are shown in the following table 2.

Table 2. Result of Text Processing

Question	What is the function of rectangular symbols in a flow chart?
Reference 1	['processing', 'data', 'process', 'calculate']
Reference 2	['show', 'processing', 'data', 'computer']
Student Answer	['indicate', 'operation', 'calculate, 'algorithm']

2.2 String Based Similarity

Automatic assessment of short answer essays in this research uses several methods including the string-based similarity method, namely the cosine coefficient and Jaccard coefficient methods [1][18] [24][36]. The process of calculating the similarity score in this study compares the calculation results of the cosine coefficient method which is calculated using (1) and the Jaccard coefficient which is calculated using (2).

$$CC_{R,S} = \frac{|R \cap S|}{|R|^{1/2} . |S|^{1/2}} \quad (1)$$

$$CC_{R,S} = \frac{|R \cap S|}{|R + S| - |R \cap S|} \quad (2)$$

To analyze the results of calculating the similarity score, this research will also carry out calculations using a modified method of the cosine coefficient and Jaccard coefficient which has been carried out in previous research by adding weighting to the word terms [1][28] namely in the form of weighted cosine coefficient (3) and weighted jaccard coefficient (4).

$$WCC_{R,S} = \frac{\sum_{i=1}^{n_r} \sum_{j=1}^{n_s} \left(W\left(R_i \cap S_j\right)\right)}{\sqrt{\sum_{i=1}^{n_r} W.R_i} . \sqrt{\sum_{j=1}^{n_s} W.Sj}} \quad (3)$$

$$WJC_{R,S} = \frac{\sum_{i=1}^{n_r} \sum_{j=1}^{n_s} \left(W\left(R_i \cap S_j\right)\right)}{\left(\sum_{i=1}^{n_r} W.R_i + \sum_{i=1}^{n_r} W.R_j\right) - \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} \left(W\left(R_i \cap S_j\right)\right)} \quad (4)$$

W is for term weight from keywords, R is for reference answers and S is for student answers. The word weighting process in previous research [1] was still carried out manually by educators using 5 categories including highly important = 5, very important = 4, important = 3, fairly important = 2, and not important = 1, whereas in this study the weighting process was based on word importance where all words contained in the 3 reference answers had a weight of 5 and for students' answers the weight was 5 if they were the same as the words in Reference answers are compared and if they are different then the default weight will be worth 1. The process of calculating the similarity score in this research was carried out in 2 stages. The similarity score calculation stage 1 is carried out to calculate the similarity score of each reference answer with the student's answer. The similarity score results obtained from the method used are then compared to select the reference answer with the highest score which is used as the answer key for the students' answers being compared, so the answer key for each student will be different according to the results of comparing the 3 reference answers that have the highest score.

2.3 Query Expansion using Bahasa Indonesia Thesaurus

The answer key that has been obtained is then compared with the student's answer. If there are the same words then the words from the answer key are stored in a new answer key and if there are different words then the word expansion process is carried out using query expansion using the Indonesian Language Thesaurus of the Department of Work Language Center. National Education 2008 as a dictionary in searching for synonyms to enrich and expand queries. If the word from the answer key has the same synonym as the word from the student's answer, then the word is changed with the synonym then saved in the new answer key and if it does not have the same synonym then no change occurs and the word is saved in the new reference answer key.

The new answer key resulting from the query expansion process is then carried out in stage 2 of the similarity score calculation process by comparing the new answer key with the students' answers to calculate the similarity score using the method used. The similarity score result of each question obtained is multiplied by the weighted value of each question given by the teacher, namely 5, then the total score of all the questions obtained is multiplied by 4 so that the test assessment results obtained by students are on a scale of 0 - 100 in accordance with the provisions of Minister of Education and Culture Regulation No. 23 2016 concerning Educational Assessment Standards

3. RESULT AND DISCUSSION

3.1 Implementation of automatic short answer scoring program

This research measures the performance of the proposed method by comparing the test scores obtained from short essay exam assessments from 2 classes, namely X Accounting and Institutional Finance (AKL) and X Office Automation and Governance (OTKP) using the cosine coefficient (CC) method. cosine coefficient using query expansion (CCQ), weighted cosine coefficient (WCC), weighted cosine coefficient using query expansion (WCCQ), jaccard coefficient (JC), jaccard coefficient using query (JCQ), and weighted jaccard coefficient (WJC), as well as weighted jaccard coefficient using query expansion (WJCQ). The comparison results of the methods used are shown in the table below.

Table 3. AKL Class Exam Results - Cosine Method

No	Teacher	CC	CCQ	WCC	WCCQ
1	76	67.30	59.05	62.48	50.40
2	82	74.61	65.51	69.58	58.41
3	68	62.48	59.23	55.49	51.00
4	88	87.66	81.74	76.32	69.52
5	60	60.83	55.82	56.24	49.18
6	76	74.50	74.50	68.38	68.38
7	80	79.28	73.60	75.57	66.79
8	84	80.26	74.42	63.30	58.01
9	84	79.94	74.12	62.78	57.62
10	80	76.22	70.40	66.90	59.62

Table 4. AKL Class Exam Results – Jaacard Method

No	Teacher	WJCQ	WJC	JCQ	JC
1	76	50.53	39.59	47.58	33.58
2	82	58.91	50.47	55.14	46.37

3	68	53.83	48.42	46.71	40.71
4	88	78.88	70.64	62.49	55.46
5	60	40.66	35.41	37.86	31.86
6	76	63.00	63.00	59.44	59.44
7	80	67.14	58.33	64.10	52.17
8	84	68.67	61.80	47.66	43.80
9	84	68.13	61.43	47.04	43.41
10	80	66.32	57.86	57.83	48.25

Table 5. OTKP Class Exam Results - Cosine Method

No	Teacher	WCCQ	WCC	CCQ	CC
1	56	52.04	47.03	50.80	43.73
2	88	84.44	76.20	83.20	71.13
3	84	80.39	75.36	70.96	65.19
4	56	58.54	47.81	50.56	40.07
5	92	91.30	88.40	88.99	84.51
6	92	89.22	84.20	85.57	77.70
7	88	81.12	75.30	67.70	62.56
8	72	72.14	61.76	55.56	44.46
9	84	76.62	68.39	71.78	59.70
10	88	87.17	84.60	81.00	77.60

Table 6. OTKP Class Exam Results - Jaccard Method

No	Teacher	WJCQ	WJC	JCQ	JC
1	56	46.12	40.90	44.29	38.29
2	88	72.80	61.84	70.96	56.96
3	84	68.90	63.92	59.20	54.52
4	56	43.97	35.37	40.36	30.86
5	92	84.32	79.87	81.40	75.40
6	92	82.28	73.23	78.92	65.92
7	88	67.86	61.16	50.10	46.47
8	72	58.78	45.50	40.0	29.18
9	84	61.24	50.30	57.19	43.19
10	88	78.40	74.17	69.42	65.14

After obtaining the test scores for class X Accounting and Financial Institutions (AKL). Testing was also carried out in class X Office Automation and Management (OTKP) from the automatic short answer scoring (ASAS) program using the proposed method and then testing the results using correlation tests and comparison tests

4. METHOD TESTING

In this article, the testing method used is to use a correlation test to determine the linear interaction or unidirectional relationship between two variables and to determine the strength of the relationship between variables. In this study, the correlation test was used to determine whether or not there was a significant relationship between the teacher's score results and the score results from the automatic short answer scoring (principal) program. The correlation test can be carried out after testing the prerequisites for data analysis, namely the normality test to find out whether the data used is normally distributed or not and the data linearity test to find out whether the data used has a significant linear relationship or not. Data analysis prerequisite testing aims to determine the type of parametric or non-parametric correlation testing that can be used to carry out correlation testing. The parametric correlation test uses the Pearson product moment correlation test for data that meets the data analysis requirements, namely normally distributed data and linear data, while the non-parametric correlation test uses the Spearman rank test for data that does not meet one or two of the data analysis requirements.

The results of prerequisite testing for data analysis on the results of grades from educators and the results of grades from the automatic essay scoring (asis) program for class normal, and the results of the teacher's scores with the results of the automatic short answer scoring (basis) program have a significant linear relationship so that the correlation test for class X Accounting and Financial Institutions (AKL) uses the Pearson product moment correlation test which is shown in the following table.

Table 7. Correlation Test Results of AKL Class using Cosine Method

Method	Pearson Correlation
Score from the teacher and score from a computer using WCCQ	0,943
Score from the teacher and score from a computer using WCC	0,883
Score from the teacher and score from a computer using CCQ	0,757
Score from the teacher and score from a computer using CC	0,723

Table 8. Correlation Test Results of AKL Class using Jaccard Method

Metode	Pearson Correlation
Score from the teacher and score from a computer using WJCQ	0,913
Score from the teacher and score from a computer using WJC	0,850
Score from the teacher and score from a computer using JCQ	0,708
Score from the teacher and score from a computer using JC	0,650

Based on the results of correlation testing in class The highest was 0.943, the result of comparing the teacher's score with the value of the automatic short answer scoring program using the weighted cosine coefficient and query expansion method. Meanwhile, using the Jaccard methods in table VIII, the highest correlation value was obtained at 0.913 as a result of comparing the teacher's scores with the automatic short answer scoring program scores using the weighted

Jaccard coefficient and query expansion. Meanwhile, using the Jaccard methods in table VIII, the highest correlation value was obtained at 0.913 as a result of comparing the teacher's scores with the automatic short answer scoring program scores using the weighted Jaccard coefficient and query expansion.

The results of prerequisite testing for data analysis on the results of grades from educators and the results of grades from the automatic essay scoring program for class, and the results of the grades from the educators with the grades from the automatic short answer scoring (basis) program have a significant linear relationship so that the correlation test for class X Office Automation and Management (OTKP) uses the Spearman rank correlation

Table 9. Correlation Test Results of OTKP Class using Cosine Method

Method	Spearmann Correlation
Score from the teacher and score from a computer using WCCQ	0,936
Score from the teacher and score from a computer using WCC	0,755
Score from the teacher and score from a computer using CCQ	0,784
Score from the teacher and score from a computer using CC	0,618

Table 10. Correlation Test Results of Class using Jaccard Method

Method	Spearmann Correlation
Score from the teacher and score from a computer using WJCQ	0,941
Score from the teacher and score from a computer using WJC	0,840
Score from the teacher and score from a computer using JCQ	0,775
Score from the teacher and score from a computer using JC	0,632

Based on the results of correlation testing in class 10 Office Automation and Management (OTKP), it shows that the results of the Spearman rank correlation test from using Jaccard methods are higher than from using cosine methods, where for using cosine methods in table ix the values obtained The highest correlation was 0.936, the result of comparing the teacher's score with the value of the automatic short answer scoring program, the weighted cosine coefficient and query expansion method, while the use of the Jaccard methods in table VIII obtained the highest correlation value of 0.941, the result of comparing the teacher's score with the value of the automatic short answer scoring program. weighted Jaccard coefficient and query expansion methods

5. CONCLUSION

Based on the test results obtained, shows that the string-based similarity method, a modification of the cosine method, namely the weighted cosine coefficient method combined with the query expansion method, has the highest correlation value in parametric correlation testing using the Pearson product-moment test in class X Accounting and Financial Institutions (AKL). amounted to 0.943, while in non-parametric correlation testing using the Spearman rank test in class X Office Automation and Governance (OTKP) it showed that the string-based similarity method was a modification of the Jaccard method, namely the weighted Jaccard coefficient method combined with the query expansion method which obtained the highest correlation value. amounted to 0.941, so the assessment results from the modified string-based similarity method combined with the query expansion method have a very strong relationship with the results from the educator assessment.

REFERENCES

[1] F. S. Pribadi, T. B. Adji, and A. E. Permanasari, "Automated Short Answer Scoring using Weighted Cosine Coefficient," *2016 IEEE Conference on e-Learning, e-Management and e-Services, IC3e 2016*. pp. 70–74, 2017, doi: 10.1109/IC3e.2016.8009042.

[2] R. Rosnelly, D. Hartama, M. Sadikin, and C. P. Lubis, "The Similarity of Essay Examination Results using Preprocessing Text Mining with Cosine Similarity and Nazief-Adriani Algorithms," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 3, pp. 1415–1422, 2021, doi: 10.17762/turcomat.v12i3.938.

[3] R. Ahmad, "E-learning Automated Essay Scoring System Menggunakan Metode Searching Text Similarity Matching Text," *J. Penelit. Enj.*, vol. 22, no. 1, pp. 38–43, 2019, doi: 10.25042/jpe.052018.07.

[4] M. J. Hazar, Z. H. Toman, and S. H. Toman, "Automated Scoring for Essay Questions in E-learning," *J. Phys. Conf. Ser.*, vol. 1294, no. 4, 2019, doi: 10.1088/1742-6596/1294/4/042014.

[5] D. Ramesh and S. K. Sanampudi, *An automated essay scoring systems: a systematic literature review*, no. 0123456789. Springer Netherlands, 2021.

[6] N. Vendyansyah and Y. A. Pranoto, "Perancangan dan Pembuatan Aplikasi untuk Mendeteksi Kemiripan Jawaban Menggunakan Cosine Similarity," *J. Tek. (Jurnal Fak. Tek. Univ. Islam Lamongan)*, vol. 13, no. 1, pp. 23–28, 2021.

[7] Febrilia E. Kurniawati and W. M. Pradnya, "Implementasi Algortma Winnowing Pada Sistem Penilaian Otomatis Jawaban Esai Pada Ujian Online

Berbasis Web," *J. Tek. Komput. AMIK BSI, Vol. VI, No.2, Juli 2020 P-ISSN 2442-2436, E-ISSN 2550-0120*, vol. VI, 2020, doi: 10.31294/jtk.v4i2.

[8] H. W. Nugroho, F. S. Pribadi, U. M. Arief, and S. Sukamta, "Perbandingan Algoritma Tf/Idf Dan Bleu Untuk Penilaian Jawaban Esai Otomatis," *Edu Komputika J.*, vol. 1, no. 2, pp. 43–51, 2014, doi: 10.15294/edukomputika.v1i2.7798.

[9] M. W. Arbiantono, "Pengembangan Aplikasi Assessment Tool Menggunakan Metode Cosine Semantic Similarity untuk Automatic Scoring Jawaban Tes Uraian pada Mata Pelajaran Basis Data di SMKN 1 Surabaya PENGEMBANGAN APLIKASI ASSESSMENT TOOL MENGGUNAKAN METODE COSINE SEMANTIC SIMIL," *Inf. Technol. Educ.*, vol. 5, no. 2, pp. 657–666, 2021.

[10] J. Zeniarta, A. Salam, and I. Achsamu, "Sistem Koreksi Jawaban Esai Otomatis (E-Valuation) dengan Vector Space Model pada Computer Based Test (CBT)," *Semin. Nas. Din. Inform.*, pp. 91–96, 2020.

[11] C. G. Abimanyu and N. A. Sanjaya ER, "Automatic Essay Answer Rating Using the Text Similarity Method," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 8, no. 4, p. 463, 2020, doi: 10.24843/jlk.2020.v08.i04.p12.

[12] E. L. Amalia, A. J. Jumadi, I. A. Mashudi, and D. W. Wibowo, "Analisis Metode Cosine Similarity Pada Aplikasi Ujian Online Otomatis (Studi Kasus JTI POLINEMA)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, p. 343, 2021, doi: 10.25126/jtiik.2021824356.

[13] R. Ahmad and R. R. O. Sasue, "Sistem Penilaian Esai Otomatis Menggunakan Algoritma Stemming Nazief dan Adriani," *J. Teknol. Transp. dan Logistik*, vol. 1, no. 2, pp. 101–108, 2020, [Online]. Available: https://jurnal.poltradabali.ac.id/jttl/article/view/15.

[14] Imam Mufiid, S. Lestanti, and N. Kholila, "Aplikasi Penilaian Jawaban Esai Otomatis Menggunakan Metode Synonym Recognition Dan Cosine Similarity Berbasis Web," *J. Mnemon.*, vol. 4, no. 2, pp. 31–37, 2021, doi: 10.36040/mnemonic.v4i2.4067.

[15] A. Condor, M. Litster, and Z. Pardos, "Automatic short answer grading with SBERT on out-of-sample questions," *Proc. 14th Int. Conf. Educ. Data Min.*, pp. 345–352, 2021.

[16] Munir, L. Septem Riza, and A. Mulyadi, "An Automatic Scoring System for Essay by Using Methods Combination of Term Frequency and n-Gram," *Int. J. Trend Res. Dev.*, vol. 3, no. 6, pp. 403–407, 2016, [Online]. Available: www.ijtrd.com.

[17] K. Srivastava, N. Dhanda, and A. Shrivastava, "An Analysis of Automated Essay Grading Systems," *Int. J. Recent Technol. Eng.*, vol. 8, no. 6, pp. 5438–5441, 2020, doi: 10.35940/ijrte.f9938.038620.

[18] T. Wahyuningsih, Henderi, and Winarno, "Text Mining an Automatic Short Answer Grading (ASAG), Comparison of Three Methods of Cosine Similarity, Jaccard Similarity and Dice's Coefficient," *J. Appl. Data Sci.*, vol. 2, no. 2, pp. 45–54, 2021, doi: 10.47738/jads.v2i2.31.

[19] S. Hassan, A. A. Fahmy, and M. El-Ramly, "Automatic short answer scoring based on paragraph embeddings," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 10, pp. 397–402, 2018, doi: 10.14569/IJACSA.2018.091048.

[20] J. Lun, J. Zhu, Y. Tang, and M. Yang, "Multiple data augmentation strategies for improving performance on automatic short answer scoring," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 13446–13453, 2020, doi: 10.1609/aaai.v34i09.7062.

[21] C. T. Lim, C. H. Bong, W. S. Wong, and N. K. Lee, "A comprehensive review of automated essay scoring (Aes) research and development," *Pertanika J. Sci. Technol.*, vol. 29, no. 3, pp. 1875–1899, 2021, doi: 10.47836/pjst.29.3.27.

[22] F. F. Lubis *et al.*, "Automated Short-Answer Grading using Semantic Similarity based on Word Embedding," *Int. J. Technol.*, vol. 12, no. 3, pp. 571–581, 2021, doi: 10.14716/ijtech.v12i3.4651.

[23] M. Afzali and S. Kumar, "Comparative Analysis of Various Similarity Measures for Finding Similarity of Two Documents," *Int. J. Database Theory Appl.*, vol. 10, no. 2, pp. 23–30, 2017, doi: 10.14257/ijdta.2017.10.2.02.

[24] U. Hasanah, A. E. Permanasari, S. S. Kusumawardani, and F. S. Pribadi, "A scoring rubric for automatic short answer grading system," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 17, no. 2, pp. 763–770, 2019, doi: 10.12928/TELKOMNIKA.V17I2.11785.

[25] F. S. Pribadi, T. B. Adji, A. E. Permanasari, A. Mulwinda, and A. B. Utomo, "Automatic short

answer scoring using words overlapping methods," *AIP Conference Proceedings*, vol. 1818. 2017, doi: 10.1063/1.4976906.

[26] H. Zhang and D. Litman, "Automated Topical Component extraction using neural network attention scores from source-based essay scoring," *arXiv*, 2020, doi: 10.18653/v1/2020.acl-main.759.

[27] E. Y. Ince and A. Kutlu, "Web-Based Turkish Automatic Short-Answer Grading System," *Nat. Lang. Process. Res.*, vol. 1, no. 3–4, p. 46, 2021, doi: 10.2991/nlpr.d.210212.001.

[28] F. Noorbehbahani and A. A. Kardan, "The automatic assessment of free text answers using a modified BLEU algorithm," *Comput. Educ.*, vol. 56, no. 2, pp. 337–345, 2011, doi: 10.1016/j.compedu.2010.07.013.

[29] U. Hasanah and D. A. Muatiara, "Perbandingan metode cosine similarity dan jaccard similarity untuk penilaian otomatis jawaban pendek," *Semin. Nas. Sist. Inf. dan Tek. Inform.*, no. 2019: SENSITIF 2019, pp. 1255–1263, 2019, [Online]. Available: https://ejurnal.dipanegara.ac.id/index.php/sensitif/article/view/511.

[30] M. D. Laxmi, Indriati, and M. A. Fauzi, "Query Expansion Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Dengan Metode Pembobotan TF-IDF Dan Algoritme Cosine Similarity Berbasis Wordnet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 823–830, 2019.

[31] F. P. Wardani and B. Rahayudi, "Query Expansion Pada Sistem Temu Kembali Informasi Dokumen Jurnal Berbahasa Indonesia Menggunakan Metode BM25," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2619–2625, 2019.

[32] H. Hariyanto, "Query Expansion Using Thesaurus for Indonesian Documents Retrieval (Temu Kembali Dokumen Berbahasa Indonesia Dengan Memperkaya Query Menggunakan Persamaan Kata Pada Tesaurus)," 2016, doi: 10.13140/RG.2.2.21986.12485.

[33] F. N. Putra, A. Effendi, and A. Z. Arifin, "Pembobotan Kata pada Query Expansion dengan Tesaurus dalam Pencarian Dokumen Bahasa Indonesia," *J. Linguist. komputasional*, vol. 1, no. 1, pp. 17–22, 2018.

[34] H. Arfandy and I. A. Musdar, "Rancang Bangun Sistem Cerdas Pemberian Nilai Otomatis Untuk Ujian Essai Menggunakan Algoritma Cosine Similarity," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 123, 2020, doi: 10.35585/inspir.v10i2.2580.

[35] M. Jamaluddin, N. Yuniarti, A. Rahmani, and J. Hutahaean, "Aplikasi Penilaian Otomatis Ujian Esai Berbahasa Indonesia Menggunakan Algoritma K-Nearest Neighbor (Studi kasus MAN Cimahi)," *Pros. Ind. Res. Work. Natl. Semin.*, vol. 10, no. August 2019, pp. 314–324, 2020, doi: 10.35313/irwns.v10i1.1404.

[36] L. Zahrotun, "Comparison Jaccard similarity, Cosine Similarity and Combined Both of the Data Clustering With Shared Nearest Neighbor Method," *Comput. Eng. Appl. J.*, vol. 5, no. 1, pp. 11–18, 2016, doi: 10.18495/comengapp.v5i1.160.

[37] S. Sugiyamto, B. Surarso, and A. Sugiharto, "Analisa Performa Metode Cosine Dan Jacard Pada Pengujian Kesamaan Dokumen," *J. Masy. Inform.*, vol. 5, no. 10, 2014, doi: 10.14710/jmasif.5.10.1-8.