

A scoring rubric for automatic short answer grading system

Uswatun Hasanah^{*1}, Adhistya Erna Permanasari², Sri Suning Kusumawardani³,
Feddy Setio Pribadi⁴

¹Department of Informatics Engineering, STMIK Amikom Purwokerto, Indonesia

^{2,3,4}Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada,
Yogyakarta, Indonesia

*Corresponding author, e-mail: uswatun_hasanah@amikompurwokerto.ac.id¹, adhistya@ugm.ac.id²,
suning@ugm.ac.id³, feddy.setio.p@mail.ugm.ac.id⁴

Abstract

During the past decades, researches about automatic grading have become an interesting issue. These studies focuses on how to make machines are able to help human on assessing students' learning outcomes. Automatic grading enables teachers to assess student's answers with more objective, consistent, and faster. Especially for essay model, it has two different types, i.e. long essay and short answer. Almost of the previous researches merely developed automatic essay grading (AEG) instead of automatic short answer grading (ASAG). This study aims to assess the sentence similarity of short answer to the questions and answers in Indonesian without any language semantic's tool. This research uses pre-processing steps consisting of case folding, tokenization, stemming, and stopword removal. The proposed approach is a scoring rubric obtained by measuring the similarity of sentences using the string-based similarity methods and the keyword matching process. The dataset used in this study consists of 7 questions, 34 alternative reference answers and 224 student's answers. The experiment results show that the proposed approach is able to achieve a correlation value between 0.65419 up to 0.66383 at Pearson's correlation, with Mean Absolute Error (MAE) value about 0.94994 until 1.24295. The proposed approach also leverages the correlation value and decreases the error value in each method.

Keywords: automatic grading, keyword matching, short answer, string-based similarity

Copyright © 2019 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

In the education field, assessment of student learning outcomes is used to measure students' ability to absorb and understand the learning process. Assessment of learning outcomes can be done with different types of questions and assessment models. The most commonly assesment methods used are multiple choice, short answer and essay [1]. As the development of information technology, assessment of learning outcomes began by utilizing e-Learning technology. This process allows the assessment of student learning outcomes automatically, using the automatic grading system. Automatic grading system has several advantages, such as being able to assess answers more objective, consistent, and faster. In addition, humans can be wrong when grading, and consistency is needed when inter-rater agreement is imperfect that may consequence from fatigue, bias, or the effects of ordering [2]. The automated scoring system for the essay can be divided into two parts, namely Automatic Essay Grading (AEG) for long answers and Automatic Short Answer Grading (ASAG) for short answers. The differences that underlie ASAG and AEG are on the length of the answers and the content being assessed. The length of short answer ranges from one phrase to a paragraph [1] and in another reference states that the length is between one phrase to four sentences [3].

Although it has several advantages, two problems also arise: the quality of the assessment and the tools used to process the language. Evaluation of the success of ASAG system is done by comparing the value generated by ASAG and the value produced by human (human rater). The value generated by ASAG must be the same or close to the direct value generated by the teacher. The value is defined as a human-computer agreement (HCA). A good HCA score is 0.75 or more [4]. Research on AEG has been widely used, but only a few ASAG studies have been done by previous researchers [5]. English became the most widely used

language in previous ASAG studies. On the other hand, the dataset is also largely private. The second problem with ASAG is about the availability of natural language processing tools. In previous studies, many researchers used the word semantic network tool, for example, WordNet. Meanwhile, WordNet for several languages is not yet available, such as for Indonesian.

In some previous studies, Latent Semantic Analysis (LSA) became a fairly common method used in an automated scoring systems [6–11]. For short answers, this method has a low performance because key term answers may only appear once or even not appear at all. The next method that is often used is from the group of String-Based Similarity method [12–16]. This methods calculates the similarity in the character (Character-Based Similarity) and the similarity in the term (Term-Based Similarity). However, this method has a weakness since the characters and terms in the answer should be exactly same as the key answer that is compared.

Burrows et al. [1] divided ASAG into five groups based on the technique used, i.e. concept mapping, information extraction, corpus-based, machine learning, and evaluation era. Roy et al. [17] divided the ASAG technique into 5 main groups, namely Natural Language Processing (NLP), Information Extraction and Pattern Matching, Machine Learning, Document Similarity, and Clustering. Each of these methods is used and grouped according to how researcher sees a problem they encounters in developing the ASAG system. Research on automatic assessment system for Indonesian language that has been done more focused on long essay answers. Some of the methods used are Cosine Similarity [13,14,18], Latent Semantic Analysis (LSA) [8, 9, 11], BLEU [19], Winnowing Algorithm [20], SVM-LSA [7], Latent Semantic Indexing (LSI) [21], and Rabin Karp's Algorithm [22]. In short answer case, some methods do not give a higher score for keywords that may appear only once or even not appear at all in the sentence.

In previous studies the existence of key terms in the answers was scored on the level of importance in the sentence [12, 23]. However, the process of scoring on each of these terms is still done manually. Some previous studies also used the corpus to facilitate variations in student answers. Somehow, a study conducted by Mohler [5] concluded that the use of large general corpus has a lower performance than a smaller domain corpus. Therefore, the next studies began to use alternative reference answers to address variations in student answers [23, 24]. This research will propose a method that is capable of handling measurements of short sentence similarities by developing approach that can be used on any domain without using word semantic tools

2. Research Method

2.1. Proposed Method

Figure 1 depicts the research flow in this work. In data collection, we use data from questions and answers on the subjects of Basic Programming Test in grade X TKJ in SMKN 8 Semarang. Secondly, pre-processing step is conducted to clean the answers from noisiness. The third step is calculating similarity and keyword matching score to get the final score. Finally, the performance evaluation is used to measure the level of agreement or value of proximity between the score generated by the system and the score generated by the teacher.

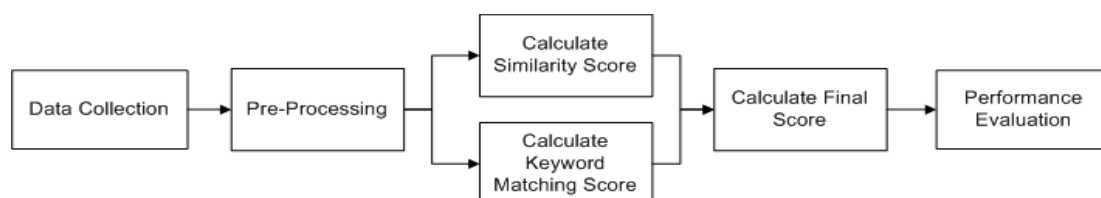


Figure 1. Research flow

2.2. Dataset

The data is in Indonesian and consists of 7 questions with 34 teacher responses as reference answers and 256 student answers in total. Each question was answered by

32 students. Table 1 shows an example of a dataset fragment consisting of questions, teacher answers and student answers.

Table 1. Example of the Dataset Fragment

Question:	<i>Apa yang kalian ketahui tentang algoritma?</i>	Score
Reference 1:	<i>Algoritma adalah urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis.</i>	4
Reference 2:	<i>Algoritma adalah suatu urutan dari beberapa langkah yang logis guna menyelesaikan masalah</i>	4
Reference 3:	<i>Algoritma adalah langkah-langkah yang disusun secara tertulis dan berurutan untuk menyelesaikan suatu masalah</i>	4
Reference 4:	<i>Algoritma adalah langkah-langkah menyelesaikan masalah secara sistematis dan logis</i>	4
Student Answers 1:	<i>langkah langkah yang logis untuk menyelesaikan masalah secara sistematis</i>	4
Student Answers 2:	<i>algoritma adalah langkah langkah yang disusun secara tertulis dan berurutan untuk menyelesaikan suatu masalah</i>	3.5
...
Student Answers 32:	<i>algoritma merupakan suatu urutan langkah langkah untuk menyelesaikan suatu masalah dalam bentuk yang logis .</i>	4

2.3. Text Pre-processing

In this study we use some text pre-processing techniques to extract the features of text which will be explained as follows:

1. Case folding is the process of converting all the characters into lowercase letters [25].
2. Tokenization is the process of converting a text into a list of words or tokens [26]. A token is a technical name for a sequence of characters [27].
3. Stemming is a technique to find the base (stem) of a word by removing its affixes [28]. The stemming process eliminates the affixes to the words so they can represent the same meaning even if they have different morphologies.
4. Stopword removal is a technique to remove words that are included in the stopwords list. The stopwords list contains a list of words that have a high number of occurrences but are less meaningful.

2.4. String-based Similarity Methods

Several methods of String-Based similarity that will be compared in this study are LCS, Cosine Coefficient, Jaccard Coefficient, and Dice Coefficient. The Longest Common Subsequence method belongs to the character-based similarity, which works by finding the same longest subsequence for all sequences in the sequence set. The calculation of LCS similarity values between sentence 1 (s_1) and sentence 2 (s_2) can be shown in (1).

$$sim_{lcs} = \frac{2 \times |LCS(s_1, s_2)|}{|s_1| + |s_2|} \quad (1)$$

Cosine Similarity is a measure of the similarity between two vectors in a dimensional space derived from the angular cosine of the multiplication of two vectors compared [29]. Jaccard Index/Jaccard Coefficient is calculated as a number of the same terms of a number of unique terms of two strings [30,31]. While Dice Coefficient is defined as twice the sum of the same terms on the two strings that are compared, divided by the total number of terms on both strings [32]. Cosine Coefficient (CC), Jaccard Coefficient (JC), Dice Coefficient (DC) [33] between document 1 (A) and document 2 (B) are defined in (3), (4) below:

$$CC(A, B) = \frac{|A \cap B|}{\sqrt{\frac{1}{|A|^2} \cdot \frac{1}{|B|^2}}} \quad (2)$$

$$JC(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (3)$$

$$DC(A, B) = 2 \frac{|A \cap B|}{|A| + |B|} \quad (4)$$

The Cosine Coefficient, Jaccard Coefficient, and Dice Coefficient Method are often used on ASAG systems and have good performance, but the methods only judge by the same word appearance without considering the order of the words.

2.5. Keyword Matching Process

At this stage we have to extract the keyword from the student's answer and the alternative answers, through the pre-processing stage. Implementation of keyword matching process between keyword provided (A) and student's answer (B), can be described as follows:

A = 'algoritma', 'langkah', 'selesai', 'masalah', 'cara', 'sistematis', 'logis'

B = algoritmaurut langkah logis selesai masalah

Number of matching keywords = {algoritma, urut, langkah, logis, selesai, masalah} = 5

Total number of keywords provided = 7

The keyword list can consist of several keywords extracted from each of the alternative answers provided. Each keyword from the alternative answer is matched to the student's answer. Furthermore, the value of the highest match score is used as the result of the reference value of the second scoring rubric.

2.6. Experimental Approach

This research is conducted through three experimental approaches. Each approach can be explained as follows:

1. The first approach: In this approach we only use one teacher's answer as key answer that was previously determined by the teacher.
2. The second approach: The second approach is done by comparing student answers with some teacher answers. In this process, the teacher provides several alternative answers.
3. The third approach: The third approach is a proposed combination technique. This approach is done by two scoring rubrics, as follows: first scoring rubric is to compare students' answers to some teacher answers (which is the second approach) and the second scoring rubric is an assessment based on keyword matching on student answers. This combination will give the final score composition as shown in Table 2. Several methods of text similarity that will be compared in this study are LCS, Cosine Coefficient, Jaccard Coefficient, and Dice Coefficient.

Table 2. Scoring Rubric

No.	Scoring Rubrics	Coefficient of Similarity/Matched Keywords	Score
1	Assessment using sentence similarity method	Range of 0 to 1	4
2	Assessment using keyword matching	Range of 0 to 1	4
Maximum score $\left(\frac{\text{Rubric 1} + \text{Rubric 2}}{2}\right)$			4

2.7. Performance Evaluation

In this study, the evaluation metrics used is Pearson correlation test. Correlation test is used to measure the level of agreement or value of proximity between the score generated by the teacher (X) and the score generated by the system (Y). The correlation value (r) is obtained from (5):

$$\text{Correlation}(X, Y) = \frac{\text{covariance}(X, Y)}{\text{standardDev}(X) \times \text{standardDev}(Y)} \quad (5)$$

In addition, *Mean Absolute Error* (MAE) is also used to measure the error rate by finding the absolute difference between the score generated by the teacher (X) and the score generated by the system (Y) and dividing it by the amount of student's answers (n). MAE is obtained by using the (6):

$$MAE = \frac{\sum |X - Y|}{n} \quad (6)$$

The expected success criteria in an automatic grading system based on the correlation value is Excellent ($r > 0.75$), Good ($r = 0.40 - 0.75$) or Poor ($r < 0.4$) [4].

3. Results and Analysis

3.1. Making Alternative Reference Answers and Pre-Processing Data

At this stage, the teacher determines alternative answers that are considered capable of representing the expected answers to the questions asked. Furthermore, keyword extraction on alternative reference answers is obtained by using text pre-processing techniques. After doing case folding and punctuation removal, we break the phrase into tokens and perform the stemming process using Sastrawi libraries (<https://github.com/sastrawi/sastrawi>) based on the Nazief-Adriani Algorithm [34], as well as the algorithm of research conducted by Asian [35], Arifin [36], and Tahitoe [37]. Lastly, we remove stopwords by using the list of ID-Stopwords libraries consisting of 758 words based on research conducted by Tala [38].

3.2. Implementation

The methods we use to assess text similarity and keyword matching are implemented using the Python programming language using the Spyder IDE. After going through the pre-processing stage, the text similarity method is used to obtain similarity values, which will be referred to as the scoring rubric 1. The methods used are Longest Common Subsequence (LCS), Cosine Coefficient (CC), Jaccard Coefficient (JC), and Dice Coefficient (DC). The next step is to do keyword matching process on student answers. The highest keyword matching value in each answer will be referred to as the scoring rubric 2. Furthermore, the final score is obtained by (7), by entering the value of the similarity score ($Score_{sim}$) and the score from the matching keyword ($Score_{keymatch}$). Each of these scores is multiplied by the *answer score*, which is 4. Note that the *answer score* may differ depending on the teacher's needs.

$$Score = \frac{(Score_{sim} * answer\ score) + (Score_{keymatch} * answer\ score)}{2} \quad (7)$$

Here is an example to calculate the similarity between answers:

Teacher's Answer = *algoritma urut langkah logis selesai masalah susun cara sistematis*

Student's Answer = *langkah logis selesai masalah cara sistematis*

$$Sim_{lcs} = \frac{2 \times 40}{58 + 40} = 0,81633 \quad Sim_{cosine} = \frac{6}{\sqrt{9} \times \sqrt{6}} = 0,81650 \quad Score_{keymatch} = \frac{6}{9} = 0,66667$$

$$Sim_{jaccard} = \frac{6}{9} = 0,66667 \quad Sim_{dice} = \frac{2 \times 6}{9 + 6} = 0,80000$$

The scoring rubric is derived from the average value between the similarity values using the String-Based method and the keyword matching value. Example of a final score calculation of the two assessment rubrics on item number 1 and student answer number 1 for LCS method can be described as follows:

$$Score = \frac{(0.81633 \times 4) + (0.85714 \times 4)}{2} = 3.34695$$

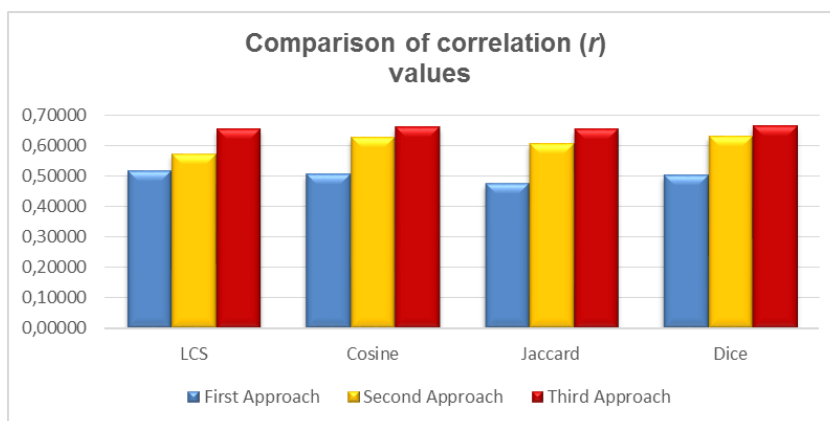
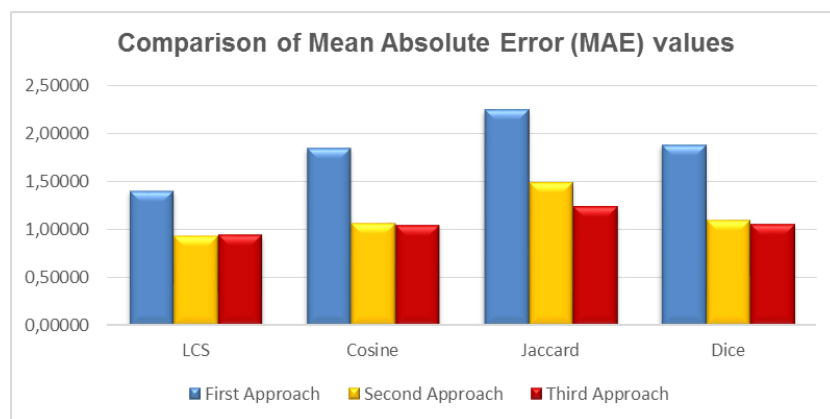
It should be noted that we get the $Score_{keymatch} = 0.66667$ in the previous example, but actually we have some references and we will take the highest $Score_{keymatch}$, that is 0.85714.

3.3. Testing

In this section we will focus on the correlation test between the score generated by the system and the score generated by the teacher. Table 3 shows the correlation (r) and Mean Absolute Error (MAE) values in the first, second, and third approaches for each method. Figure 2 and Figure 3 shows a summary of the first, second, and third approaches, by comparing the correlation and MAE values between the methods.

Table 3. Results of Correlation and *MAE* Values for each Approaches

Approaches	LCS		Cosine		Jaccard		Dice	
	<i>r</i>	<i>MAE</i>	<i>r</i>	<i>MAE</i>	<i>r</i>	<i>MAE</i>	<i>r</i>	<i>MAE</i>
1 st	0.51681	1.39910	0.50700	1.85451	0.47411	2.25458	0.50334	1.88734
2 nd	0.57050	0.93334	0.62644	1.07185	0.60442	1.49105	0.63070	1.10068
3 rd	0.65419	0.94994	0.66019	1.04087	0.65517	1.24295	0.66383	1.05381

Figure 2. Comparison of correlation (*r*) values between methodsFigure 3. Comparison of *MAE* values between methods

3.4. Discussion

The third approach, which is a combination of two assessment rubrics, can enhance the correlation value and decrease the *MAE* score. Dice Coefficient method has the highest correlation on the total value of 0.66383, but has a large *MAE* value that is 1.05381. The Cosine Coefficient method can be considered to have a correlation value approaching the Dice Coefficient method (0.66019), with smaller *MAE* values (1.04087). The LCS method has the smallest *MAE* value (0.94994) with a correlation value of 0.65419. The type of question used in this study is to mention the definition and the steps. This type of question tends to have a more structured answer, using active sentences.

The study also observed that there are two different ways when students fill in the answers. First, the students mention it completely by referring to the subject of what is asked in the question. Secondly, students only answer what is asked without rewriting the subject of question. The completeness of the sentence has certainly affected the value of the similarity of the answer, especially if there is only one reference. To overcome this, teachers can create an

answer model by providing a predefined format. In addition to using the above methods, the use of alternative answers is also able to overcome the problem of variation in student answers. However this alternative answer should be handled automatically for the assessment efficiency.

4. Conclusion

The proposed approach is able to work well in handling sentence similarities in Indonesian rather than conventional methods, as indicated by increasing correlation and decreasing *MAE* values. Further research is expected to be able to process the dataset with a more varied question type, which is still included in the short answer rules. Further research is also expected to generate an alternative answer automatically, and develop algorithms that can extract keywords automatically. In order to minimize errors when filling in answers, a spell corrector device can be added. Limitations of sentence length can also be implemented so that students only focus with the questions asked.

References

- [1] Burrows S, Gurevych I, Stein B. The Eras and Trends of Automatic Short Answer Grading. *Int J Artif Intell Educ.* 2015; 25(1): 60–117.
- [2] Haley DT, Thomas P, De Roeck A, Petre M. *Measuring Improvement in Latent Semantic Analysis-Based Marking Systems: Using A Computer to Mark Questions About HTML.* Conf Res Pract Inf Technol Ser. 2007; 66: 35–42.
- [3] Siddiqi R, Harrison CJ, Siddiqi R. Improving Teaching and Learning Through Automated Short-Answer Marking. *IEEE Trans Learn Technol.* 2010; 3(3): 237–249.
- [4] Fleiss J, Levin B, Cho Paik M. *Statistical Methods for Rates and Proportions.* Third Edit. Technometrics. 2004; 46: 263-264.
- [5] Mohler M, Mihalcea R. *Text-to-text Semantic Similarity for Automatic Short Answer Grading.* In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL '09). 2009: 567–575.
- [6] Perkasa DA, Saputra E, Fronita M. Sistem Ujian Online Essay Dengan Penilaian Menggunakan Metode Latent Semantic Analysis (LSA) (*Online Essay Examination System with Assessment Using the Latent Semantic Analysis (LSA) Method.* *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi.* 2015; 1(1): 1–9.
- [7] Adhithia R, Purwarianti A. Penilaian Esai Jawaban Bahasa Indonesia Menggunakan Metode Svm - Lsa Dengan Fitur Generik (*Assessment of the Answer Essay in Indonesian Using the SVM - LSA Method with Generic Features*). *J Sist Inf.* 2012; 5(1): 33.
- [8] Aji RB, Baisal ZA, Firdaus Y. Automatic Essay Grading System Menggunakan Metode Latent Semantic Analysis (*Automatic Essay Grading System Using the Latent Semantic Analysis Method*). *Semin Nas Apl Teknol Inf.* 2011; 2011: 1–9.
- [9] Yustiana D. Penilaian Otomatis Terhadap Jawaban Esai Pada Soal Berbahasa Indonesia Menggunakan Latent Semantic Analysis (*Automatic Scoring of Essay Answers in Indonesian Questions Using Latent Semantic Analysis*). *Semin Nas Inovasi dalam Desain dan Teknol-IDEaTech 2015.* 2015;123–130.
- [10] Klein R, Kyrilov A, Tokman M. Automated Assessment of Short Free-Text Responses in Computer Science using Latent Semantic Analysis. In: *ITiCSE '11.* 2011: 158–62.
- [11] Agung A, Ratna P, Budiardjo B. Simple: Sistem Penilai Esei Otomatis untuk Menilai Ujian in Bahasa Indonesia (*Simple: Automatic Essay Assessment System for Assessing Exams in Indonesian*). *Makara, Teknol.* 2007; 11(1): 5–11.
- [12] Pribadi FS, Adji TB, Permanasari AE. *Automated Short Answer Scoring using Weighted Cosine Coefficient.* In: 2016 IEEE Conference on e-Learning, e-Management and e-Services, IC3e 2016. 2017: 70–74.
- [13] Lahitani AR, Permanasari AE, Setiawan NA. Cosine similarity to determine similarity measure: Study case in online essay assessment. In: *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016.* 2016: 1–6.
- [14] Fitri R, Asyikin AN. Aplikasi Penilaian Ujian Essay Otomatis Menggunakan Metode Cosine Similarity (*Automatic Essay Exam Assessment Application Using the Cosine Similarity Method*). *J POROS Tek.* 2015; 7(2): 88–94.
- [15] Rodrigues F, Araújo L. Automatic Assessment of Short Free Text Answers. In: *Proceedings of the 4th International Conference on Computer Supported Education.* 2012: 50–7.
- [16] Gomaa W, Fahmy A. Short Answer Grading Using String Similarity And Corpus-Based Similarity. *Int J Adv Comput Sci Appl.* 2012; 3(11): 115–21.

- [17] Roy S, Narahari Y, Deskhmukh OD. *A Perspective on Computer Assisted Assessment Techniques for Short Free-Text Answers*. In: International Computer Assisted Assessment Conference. 2015: 96–109.
- [18] Pramukantoro ES, Fauzi MA. *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: 149–55.
- [19] Nugroho HW, Pribadi FS, Arief UM, Sukamta S. Perbandingan Algoritma TF/IDF Dan BLEU Untuk Penilaian Jawaban Esai Otomatis (*Comparison of TF/IDF and BLEU Algorithms for Automatic Essay Answer Assessment*). *Edu Komputika J*. 2014; 1(2): 43–51.
- [20] Astutik S, Cahyani AD, Sophan MK. Sistem Penilaian Esai Otomatis Pada E-Learning Dengan Algoritma Wininging (*Automatic Essay Assessment System on E-Learning with Wininging Algorithms*). *J Inform*. 2014; 12(2): 47–52.
- [21] Riwayati I, Hartati I, Purwanto H, Suwardiyono. Penilaian Jawaban Essay Menggunakan Semi Discrete Decomposition Pada Metode Latent Semantic Indexing (Essay Answer Assessment Using Semi Discrete Decomposition in the Latent Semantic Indexing Method). In: Seminar Nasional Aplikasi Sains & Teknologi (SNAST). 2014: 211–216.
- [22] Bahri S. Penilaian Otomatis Ujian Essay Online berbasis Algoritma Rabin Karp. Swabumi. 2014; 1(September).
- [23] Noorbehbahani F, Kardan AA. The Automatic Assessment of Free Text Answers Using a Modified BLEU Algorithm. *Comput Educ*. 2011; 56(2): 337–45.
- [24] Omran AM Ben, Ab Aziz MJ. Automatic Essay Grading System for Short Answers in English Language. *J Comput Sci*. 2013; 9(10): 1369–1382.
- [25] Christopher DM, Prabhakar R, Hinrich S. Introduction to Information Retrieval. Introduction to information retrieval. Cambridge University Press. 2008; 1: 496.
- [26] Perera GR, Perera DN, Weerasinghe a. R. A Dynamic Semantic Space Modelling Approach for Short Essay Grading. In: 15th International Conference on Advances in ICT for Emerging Regions, ICTer 2015-Conference Proceedings. 2016: 43–9.
- [27] Bird S, Klein E, Loper E. Natural language processing with Python: Analyzing Text with the Natural Language Toolkit. 1st ed. CA: O'Reilly Media, Inc. 2009: 7.
- [28] Carvalho G, de Matos DM, Rocio V. Document Retrieval for Question Answering. In: Proceedings of the ACM first PhD workshop in CIKM on-PIKM '07. 2007: 125.
- [29] Saputra WSJ, Muttaqin F. Pengenalan Karakter Pada Proses Digitalisasi Dokumen Menggunakan Cosine Similarity (*Character Recognition in the Process of Digitizing Documents Using Cosine Similarity*). *Semin Nas Tek Inform*. 2013: 51–56.
- [30] Jaccard P. Etude De La Distribution Florale Dans Une Portion Des Alpes Et Du Jura. *Bull la Soc Vaudoise des Sci Nat*. 1901; 37(142): 547–579.
- [31] H.Gomaa W, A. Fahmy A. A Survey of Text Similarity Approaches. *Int J Comput Appl*. 2013; 68(13): 13–18.
- [32] Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945; 26(3): 297–302.
- [33] Thada V, Jaglan V. Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm. *Int J Innov Eng Technol*. 2013; 2(4): 202–205.
- [34] Nazief B, Adriani M. Confix Stripping: Approach to Stemming Algorithm in Bahasa Indonesia. *Intern Publ Fac Comput Sci Univ Indones Depok*, Jakarta. 1996;
- [35] Asian J. Effective Techniques for Indonesian Text Retrieval. 2007.
- [36] Arifin AZ, Mahendra IPAK, Ciptaningtyas HT. Enhanced Confix Stripping Stemmer and Ants Algorithm For Classifying News Document In Indonesian Language. In: Proceeding of International Conference on Information & Communication Technology and Systems (ICTS). 2009: 149–158.
- [37] Tahitoe AD, Purwitasari D. Implementasi Modifikasi Enhanced Confix Stripping Stemmer in Bahasa Indonesia dengan Metode Corpus Based Stemming. Jurusan Teknik Informatika, Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember (ITS)–Surabaya. Surabaya. 2010.
- [38] Tala FZ. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. M.Sc. Thesis, Appendix D. Amsterdam. 2003.