

ICESTI2017AES

by Viny M

Submission date: 08-Nov-2020 02:43PM (UTC+0700)

Submission ID: 1439401676

File name: matecconf_icesti2018_01037.pdf (1,011.83K)

Word count: 6180

Character count: 31532

Automatic Essay Scoring in E-learning System Using LSA Method with N-Gram Feature for Bahasa Indonesia

Rico Setiadi Citawan*, Viny Christanti Mawardi, and Bagus Mulyawan

Faculty of Information Technology, Tarumanagara University, Letjend S. Parman No. 1, West Jakarta 11440, Indonesia

Abstract. In the world of education, e-learning system is a system that can be used to support the educational process. E-learning system is usually used by educators to learners in evaluating learning outcomes. In the process of evaluating learning outcomes in the e-learning system, the form type of exam questions that are often used are multiple choice and short stuffing. For exam questions in the form of essays are rarely used in the evaluation process of educational because of the difference in the subjectivity and time consuming in the assessment process. In this design aims to create an automatic essay scoring feature on e-learning system that can be used to support the learning process. The method used in automatic essay scoring is Latent Semantic Analysis (LSA) with n-gram feature. The evaluation results of the design features automatic essay scoring showed that the accuracy of the average achieved in the amount of 78.65 %, 58.89 %, 14.91 %, 71.37 %, 64.49 % in the LSA unigram, bigram, trigram, unigram + bigram, unigram + bigram + trigram.

Key words: Automatic essay scoring, e-learning, latent semantic analysis, n-gram, singular value decomposition

1 Introduction

Along with advancement of internet technology rapidly allows the application of technology in various fields including one in the education field. One of the technologies that can be utilized to support the learning process is e-learning system. In e-learning system many features are developed to support the learning process such as online exam features who provided by educators to learners in evaluating learning outcomes.

In e-learning system, evaluation of learning outcomes in essay form is important because evaluation results in essay form can be used by educators as an indicator to measure the level of learners understanding of the exams given. Therefore, to solve the problem is designing e-learning system with automatic essay scoring feature that can be used to support the learning process between educators and learners.

There are several methods that can be used to perform assessments in automatic essay scoring such as string matching algorithms like Rabin Karp, Knuth Morris Pratt,

* Corresponding author: rsctvx@gmail.com

Levenshtein distance, etc. However, for this designing is used method Latent Semantic Analysis (LSA) which can be used to find the similarity between answer student essays with answer key model lecturer in automatic essay scoring.

LSA is a method in information retrieval [12] that represents words along with sentences in a matrix with mathematical calculations. Mathematical calculations are performed by mapping [16] the presence or absence of words from word groups in the matrix. LSA method uses a linear algebra technique called Singular Value Decomposition (SVD) to find patterns of relationships between words and contexts that contain [10] words in semantic matrix.

Generally existing Automatic Essay Scoring (AES) techniques which are using LSA method do not consider word order in a sentence. To solve the problem, the LSA method used in automatic essay scoring for this designing will be added with n-gram feature so that word order in sentence can be considered.

4

2 Literature Review

1

In this case, the Latent Semantic Analysis (LSA) method with n-gram feature are used to create an automatic essay scoring in e-learning system. 1

4

2.1 Pre-processing

Pre-processing is an early stage in processing input data before entering the main stage process. Pre-processing consists of several stages. The pre-processing stages are performed in the system are case folding, lexical analysis, tokenization. Here is an explanation of the [4] pre-processing steps performed as follows:

i) Case Folding

Case folding is the stage that changes all the letters in the corpus into lowercase. The purpose of case folding is to uniform input data because the input can be lowercase and uppercase.

ii) Lexical Analysis

Lexical analysis is the process of eliminating numbers, punctuation and characters which cannot be read by the system.

iii) Tokenization

Tokenization is the process of cutting or separating words from a collection of text or corpus.

13

2.2 Latent semantic analysis (LSA)

LSA stands for Latent Semantic Analysis is an information retrieval method that works on a set of texts in which words can be represented along with their context in a matrix with mathematical calculations. The formed matrix is a mapping of the presence or absence of a word or term from a particular group of words or terms. In addition, the context used can be a list of keywords, sentences, even paragraphs.

Latent Semantic Analysis (LSA) was proposed and patented [19] 1988 by Scott Deerwester, Thomas Launder, and Susan Dumais [1]. This method is widely used in the field of Information Retrieval and Natural Language Processing. One example of the use of [18] LSA method is the assessment of essay answers. The LSA method uses the mathematical technique of linear algebra called Singular Value Decomposition (SVD) in forming semantic matrix.

The semantic matrix that has been formed is a matrix that contains the pattern of relationship between words with the context in matching between the answer and answer

key [2]. LSA methods generally work based on the unigram word occurring simultaneously in certain contexts so that LSA method is more concerned with the keywords contained in a sentence regardless of the word orders and grammar in a sentence, Therefore, to solve the problem, in this LSA method will be added n-gram feature so that LSA method can work on the appearance of word and phrase in certain context so that word order in sentence can be considered [3].

The designing of automatic essay scoring using LSA method, the first step is to represent the student answer and lecturer answer key in a matrix by forming term frequency matrix of student answer and term frequency matrix of lecturer answer key. The term frequency matrix is formed by counting the number of times the word appears on the lecturer answer key. The frequency of occurrence of a word or term that counts can be a set of unigram, bigram, trigram, or all of them.

The formed term frequency matrix consists of two are the term frequency matrix of the lecturer answer key and the frequency term matrix of the student answer. The frequency term matrix for the lecturer answer key consists of rows and columns where each row in the matrix represents a unique word in the lecturer answer key while each column in the matrix represents each sentence of lecturer answer key. In addition, each cell in the matrix represents the number of times the occurrence of a unique word that exists in each sentence of lecturer answer key.

In the term frequency matrix for student answers also consists of rows and columns where each row in the matrix represents a unique word on the lecturer answer key while each column in the matrix represents every student answers sentence then each cell on the matrix represents the number of occurrences of a unique word lecturer answer key is in every student answer sentence. The weighting used in each matrix cell is the frequency of the student's answer and the lecturer's answer key are the weighting of the term frequency (tf).

The next stage is to calculate Singular Value Decomposition (SVD) on the term frequency matrix that has been formed in the previous stage is the term frequency matrix for lecturer answer key and term frequency matrix for student answers. The purpose of SVD is discover a new pattern or relationship "latent" between words and sentences containing these words. The result of the SVD calculation process produces three matrices components are two orthogonal matrices and one diagonal matrix. The factoring of SVD from a matrix A dimension txd is as follows:

$$A_{t \times d} = U_{t \times n} \times S_{n \times n} \times V_{d \times n}^T \quad (1)$$

Where:

A: term matrix sized txd.

U: orthogonal matrix sized txn.

S: diagonal matrix sized nxn.

V: orthogonal matrix transpose sized dxn.

t: number of rows of the matrix A.

d: number of columns of the matrix A.

n: the rank value of the matrix A ($\leq \min(t, d)$).

The decomposed of matrix A produces three matrices of the matrix U, S, and VT. The size or dimensions of three matrices generated in the SVD process are different. In the matrix U has the same row as the matrix A, but has fewer columns than the matrix A is n column. In the matrix S has the size of n row and n columns. In the matrix VT has the same column as the matrix A, but has fewer row sizes than matrix A or equal to column matrix A is d row.

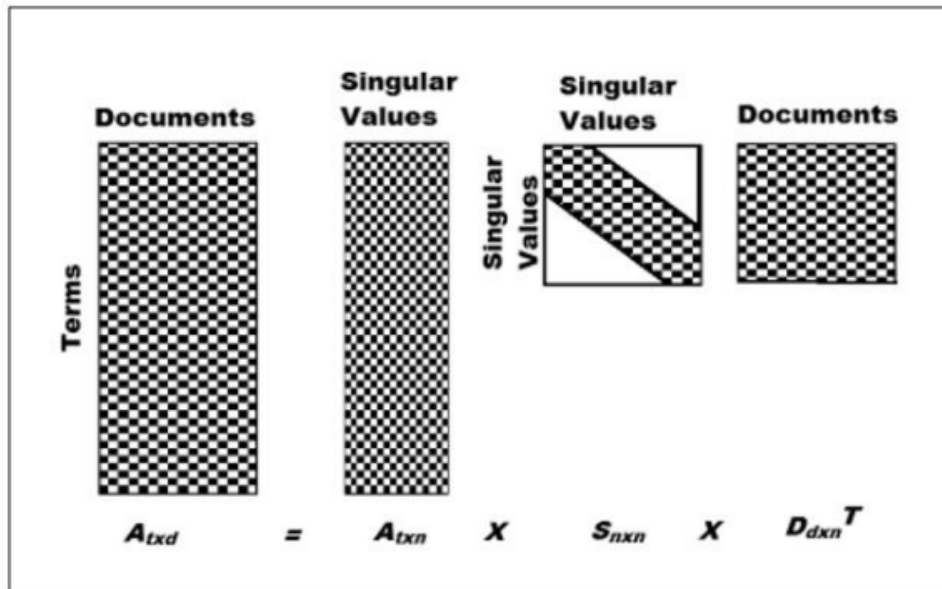


Fig. 1. Singular value decomposition of matrix [2].

Based on Figure 22, the column of the matrix U is an orthogonal eigenvector of $A \times A^T$. In column matrix V is an eigenvector of orthogonal $A^T \times A$. In the matrix S is a diagonal matrix whose value is the root of the eigenvalues of the matrix U or V . After doing the SVD process, the next step is to reduce the matrix of SVD calculation result or smoothing process. The purpose of reducing the size or dimension of the matrix is to remove noise or words that are considered unimportant without eliminating the correlation or the relationship between words and sentences [4]. The reduction of the matrix dimension is done by reducing the dimension of the diagonal matrix containing the singular value of the second matrix result of the SVD process.

When reducing the size of the diagonal matrix S containing the singular value, the first step is, choose a value of k . The value of k which can be selected in reducing the matrix size is $k \leq n$. After that the value of k will be used on the matrix of the SVD calculation results by taking the firsts row data and the first column k data on the diagonal matrix containing the singular value [5]. Unselected dimension the singular value on the diagonal matrix S will be set to zero. The reduction in the diagonal matrix (S) also affects both orthogonal matrices U and V .

The result of the reduced SVD matrices can be used for reconstructing the matrix by using a lower dimension (A_k). Reconstructed matrix (A_k) has the same size as the previous matrix (A). However, the reconstructed matrix (A_k) is a matrix with approximation of matrix A using a lower dimension which can show the hidden relationship between words in the sentence so that it can cover all important words and can ignore unimportant words that are not considered.

In addition, the reconstructed matrix can estimate the hidden new relationship of a group of words or terms in a particular sentence so that the word or term group contained in the sentence may be close to or relate to each other in space k although the word or term is not ever present or appear on the sentence. Reconstruction matrix of SVD results can be seen in the Equation 2 and Figure 2.

$$A_{t \times d} = A_k = U_k \times S_k \times V_k^T \quad (2)$$

Where:

A_k : reconstruction of term frequency matrix sized k.

U_k : orthogonal matrix sized k.

S_k : diagonal matrix sized k.

V_k : orthogonal matrix transpose sized k.

k: parameter of reduction dimension matrix.

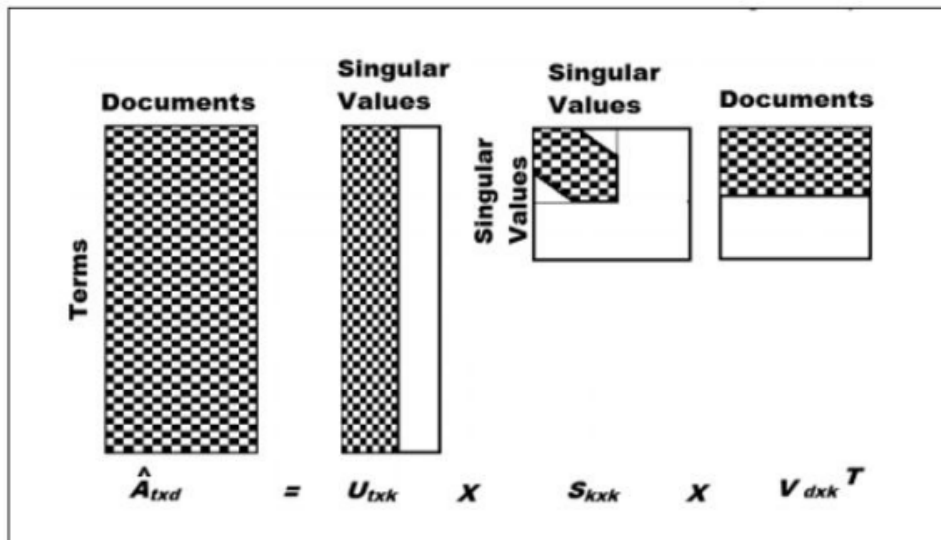


Fig. 2. Reconstruction matrix of singular value decomposition.

After the reconstruction of the student answer matrix and the lecturer answer key matrix has been formed, the next step is to form the answer vector and answer key vector of the matrix result from the student answer and the lecturer answer key who have been reconstructed again using the lower dimension. Reconstruction using the lower dimension will produce a new representation or relationship on the student answer and the lecturer answer [6]. In this case the answer vector is the student answer sentence and the answer key vector is the lecturer answer key sentence. The final stage of the automatic essay scoring using LSA is to measure the similarity between the vector of students answers and the vector of lecturer answer key.

2.3 Cosine similarity

Cosine Similarity is a technique for measuring the similarity of the angle cosine value between document vectors and query vectors. In this case the document vector represents the sentence in the student answer while the query vector represents the sentence in the lecturer answer key. After the document vector (d) and query vector (q) has been formed, the similarity between the document vector and the query vector can be calculated using the Equation (3).

$$\text{CosineSim}(q, d) = \frac{\sum_{j=1}^t w_{qj} \cdot w_{dij}}{\sqrt{\sum_{j=1}^t (w_{dij})^2} \sqrt{\sum_{j=1}^t (w_{qj})^2}} \quad (3)$$

Where:

q; the student answer vector.

d: the lecturer answer key vector.

4

2.4. N-gram

Basically, the n-gram model is a probabilistic model designed by mathematicians in the early 20th century and later developed to predict the next item in the order of items. The item can be a sequence letter/character or a word sequence [7]. In the word generation process, n-gram consists of words sequence along n words in a sentence. Where i word, n-gram is can be used to extract pieces of n words from a sentence or paragraph. A n-gram of size 1 is called unigram, for n-gram of size 2 called bigram, for n-gram of size 3 called trigram, whereas for n-gram is a term when it has a size greater than 3. For example, the word sequence "perangkat keras komputer" (computer hardware), the word sequence in n-gram as follow.

Unigram : *perangkat, keras, komputer.* (device, hard, computer)

Bigram : *perangkat keras, keras komputer.* (hard device, computer hard)

Trigram : *perangkat keras komputer.* (computer hardware)

2.5. Evaluation

Evaluation of an automated essay scoring system is needed to assess how the Latent Semantic Analysis method works good in assessing or scoring the essay answers. This evaluation is done by comparing the judgments which generated by the system with human judgment, in this case the lecturer. After all data was already obtained, the next step will be analyzing mean absolute error and mean absolute percentage error result. The result of mean absolute error is the average difference between value from lecturer and systems.

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (4)$$

Where: 5

MAE : average value from lecturer and system.

X_i : value from lecturer.

Y_i : value from system.

n : numbers of data.

5
While the result of mean absolute percentage error is the average difference between value from lecturer and systems. Then the result of mean absolute percentage error calculation can also be measured the success rate of the automatic essay scoring system in assessing or scoring student essay answers.

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right|}{n} \times 100 \quad (5)$$

Where:

MAPE : average percentage error value.

X_i : value from lecturer.

Y_i : value from system.

n : numbers of data.

3 Planning and implementation

3.1. System plan

At this stage will explain the planning that needs to be done in system plan. This stage is done by first determining the purpose and use of the system to be designed. The designed system is an e-learning system with automatic essay scoring feature. The system is built and is intended for students and lecturers in the learning process where in this system planning students can do the exam given by the lecturer in the form of essays while the lecturer can give the test to the student in the form of essay which then will be automatically assessed by the system with the answer key who has been included by lecturer.

The initial stage of the automatic essay scoring system is the preparation stage of the required data such as student answers, questions and models of lecturer answers key. Data consisting of questions and answers key are taken from the book "Discovering Computers 2010: Living in Digital World", websites, ppt and pdf about course material "introductory information technology".

Meanwhile the student answer data is collected by distributing questionnaires containing questions about course material introductory information technology to the students of the information technology faculty Tarumanagara University. In designing automatic essay scoring system that is built, to facilitate the design of the system in assessing essay answers, then the collected data used in the system such as student answers, questions and models of lecturer answer key will be stored ⁸ the database.

Automatic essay scoring system designed to receive ⁸ input in the form of questions, answer keys, and answers essay. Questions and answers were obtained from the lecturer while the essay answers were obtained from the student. Then the answers key and answers will be done pre-processing, after that the system will perform the similarity calculation between the student's essay answers with each model of lecturer's answer key using Latent Semantic Analysis (LSA) where the highest similarity values obtained using cosine similarity will be used to produce the student scores output.

In the automated essay scoring system using LSA designed as shown in Figure 3, this system runs by first managing the data entered by the user consisting of lecturers and students. Data that become system input is lecturer answer key data and student answer data. Once the system gets input from the user, the system goes into the pre-processing stage, the process of calculating the Latent Semantic Analysis algorithm, the result of the student's answer score.

Based on the flow in Figure 3, there is a pre-processing stage. Pre-processing is a series of steps performed to manage input data before entering the next stage. In the pre-processing staged there are several stages performed such as case folding, lexical analysis, and tokenization.

Further data that has been pre-processing this will be input or enter in the process of latent semantic analysis algorithm. Initial stages performed in the process of latent semantic analysis algorithm is the formation of term frequency matrix for answer and term frequency matrix for answer key. The process of forming term frequency matrix by counting the number of words or terms that appear in each sentence in the matrix, then the term frequency matrix that has been formed will be calculated SVD or Singular Value Decomposition that will produce the matrices U, S and V. The result of SVD matrices calculation will be reduction process which then be used to reconstruct the matrix by using a lower dimension.

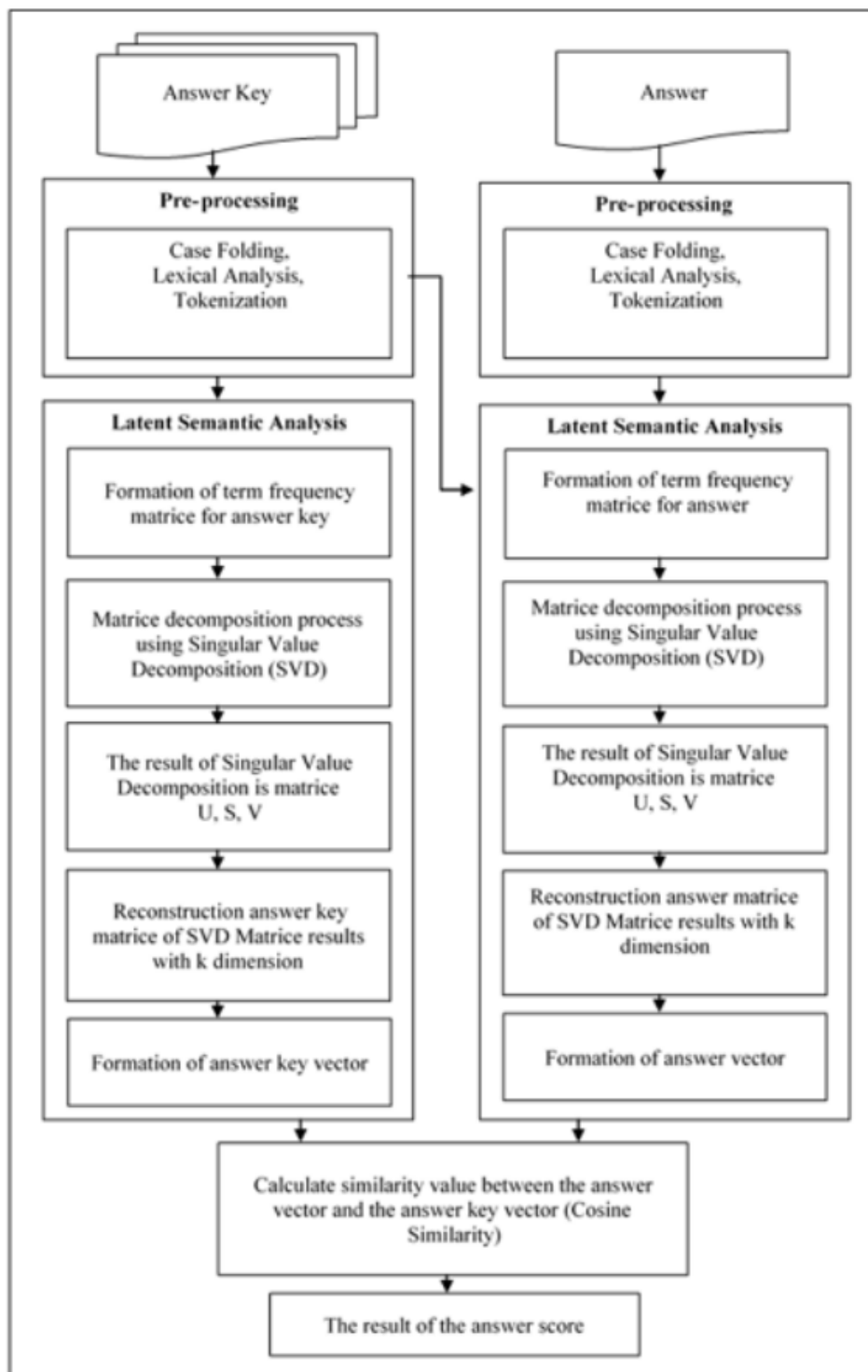


Fig. 3. Similarity answer key and answer using LSA.

In the formation of the term frequency matrix of the answer, the set of terms or words used to calculate the frequency of words occurrence in each sentence answer is a collection of terms or words derived from the pre-processing answers key. Likewise, for the process of forming the term frequency matrix of answer key. If the reconstruction answer matrix

and answer key matrix has been formed, then the next process is to form an answer vector and answer key vector.

The answer vector and the answer key vector that has been formed will be calculated by the value of similarity using Cosine Similarity method where a vector formed is a representation of a sentence in the matrix column. Then the value used to evaluate each sentence answer is the value with the highest similarity between answers vector with all the answers key vector.

In the process of assessing student answers, the value to be taken to be the output is the average value obtained from the sum of each value of the highest similarity between the student answer sentences that is compared with all key lecturer answer sentences divided by the number of sentences of student answers. Then the average value of similarity that has range (0 to 1) multiplied by value 100.

3.2. System implementation

After the design phase of the system has been designed then the next stage is to make software adjustments with the system design that has been made before. The first thing to consider in the making of this system **4** the system needs consisting of hardware and software from the server. Server serves as a mediator during the process of data transfer using internet connection, on the manufacture of the system will use the hosting server.

The hardware used for system development used a computer with IBM Server computer specifications as a server. As for the client side can use a computer that has a web browser like internet explorer, google chrome, mozilla firefox, etc. Software used on the server that is IIS manager and SQL Server Management Studio are used as central data storage server, Matlab used to perform mathematical calculations of linear algebra, and ASP.NET C# and Perl is used as a programming language (back end).

After all the hardware and software that will be used in the manufacture is available, then the creation of the system can be started by first setting up the database needed to store data before and after processing. The existing database will be created using the SQL Server database server.

After the database is created, then the next stage to do is create interface display for each module that has been designed previously. Display interface is made as simple as possible so that users can use the existing program easily. Once the existing interface is created, it will make the program code for the existing interface so that the interface is made to work properly.

If the existing program code has been created, then will be tested against each module that has been made. Testing is made to be known whether the existing program has been running as expected or not. In addition, through the tests made can also be known whether the program is made there is still an error or not.

Once the interface has been completed, program coding will be made available for automatic essay scoring features with the algorithm and the theoretical base described earlier. When the coding program has been completed, it will be tested against the features that have been made. Feature testing is done to find out whether the program has been made output according to the designed.

4 Result and discussion

Testing conducted consist of two stages is black box testing and automatic essay scoring features testing. Black box testing is done to test whether each function in each module has been able to function properly. In testing the automatic essay scoring feature is performed

to determine whether the system created can assess the essay answers according with the judgments made by humans.

4.1. Modules testing

This testing is performed on all modules that exist on the system created. Outline, the modules that exist in the system is divided into 2 main modules namely the module for student and modules for lecturer.

4.1.1 Automatic essay scoring modules testing

The results of the testing conducted found that all the functions of the existing automatic essay scoring module had been able to run according with its function. Here is a picture showing the test results from one of the automatic essay scoring modules.

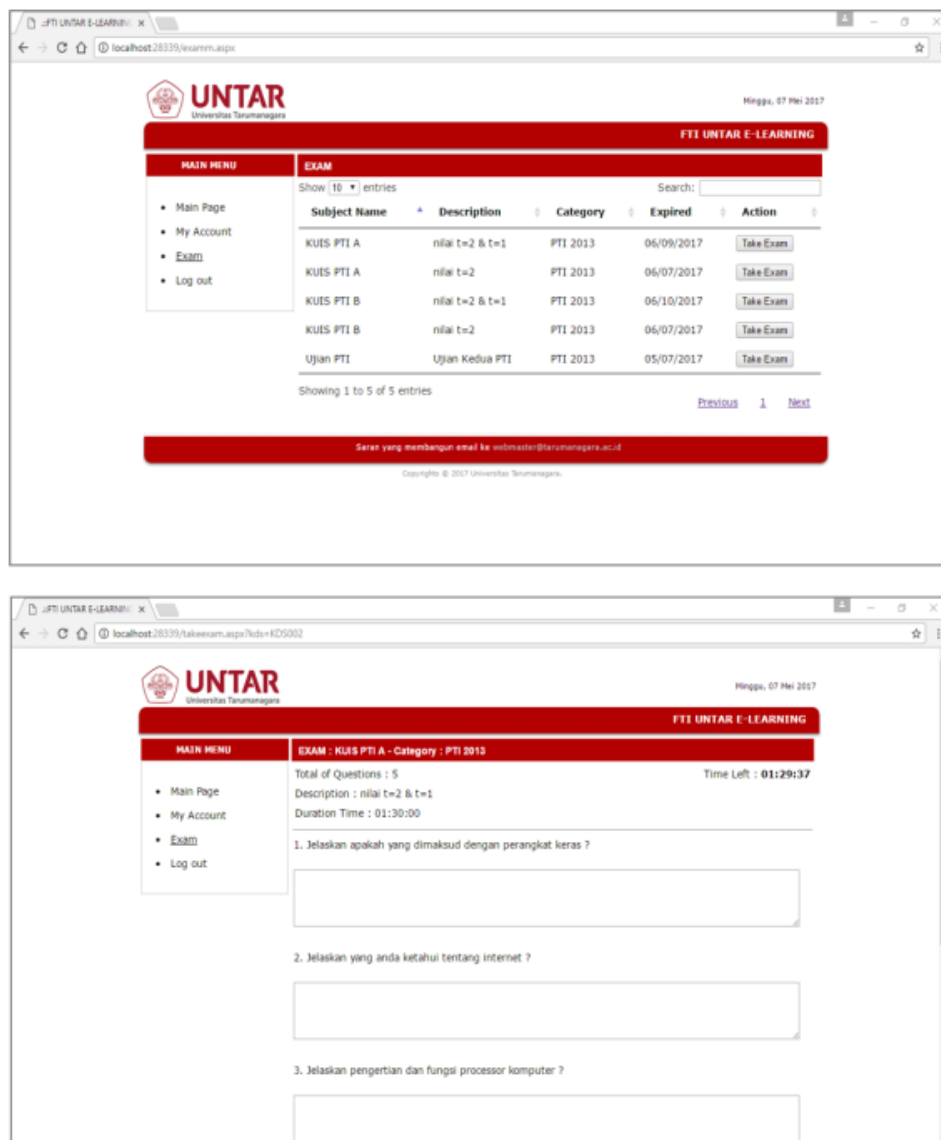
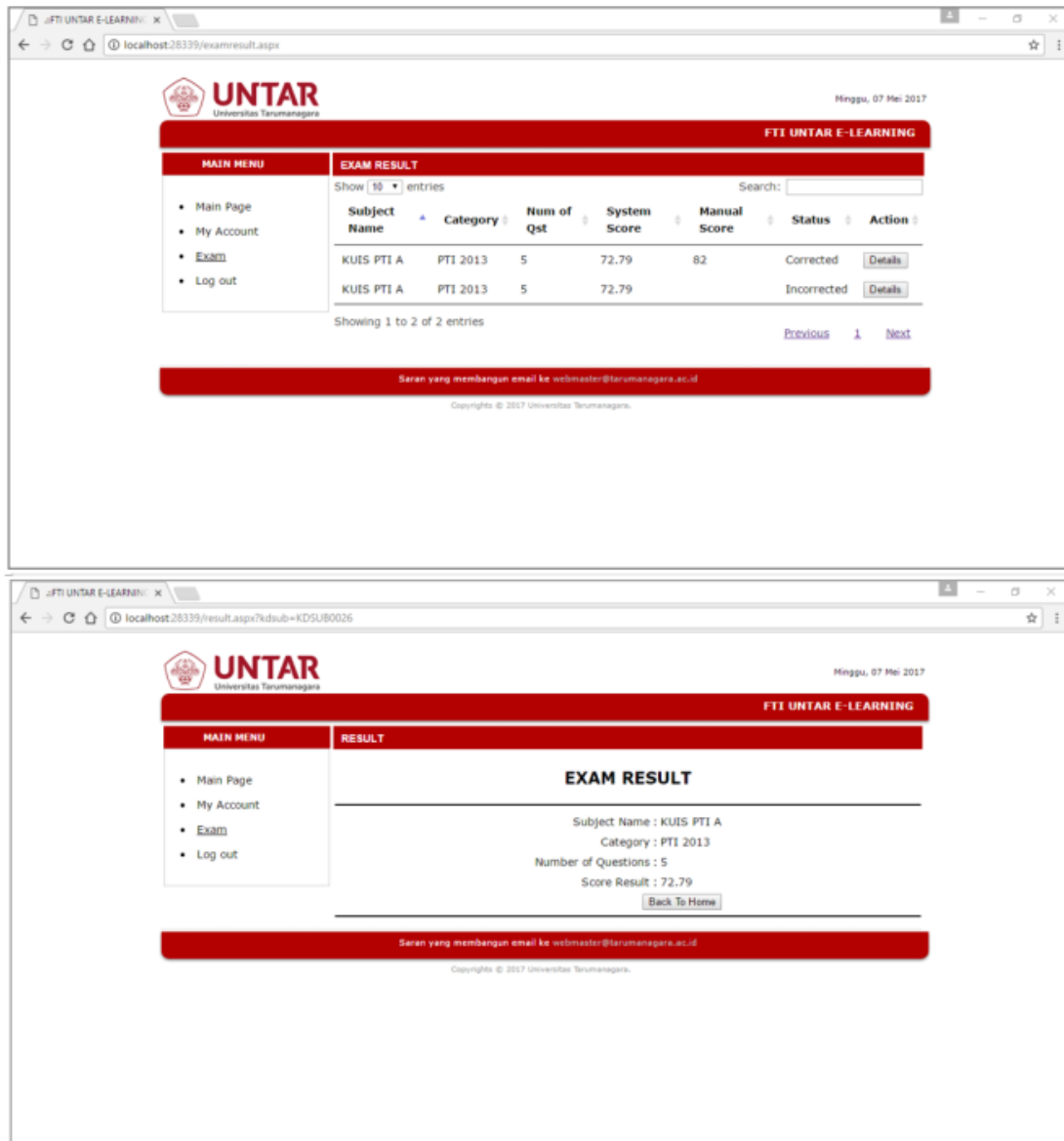
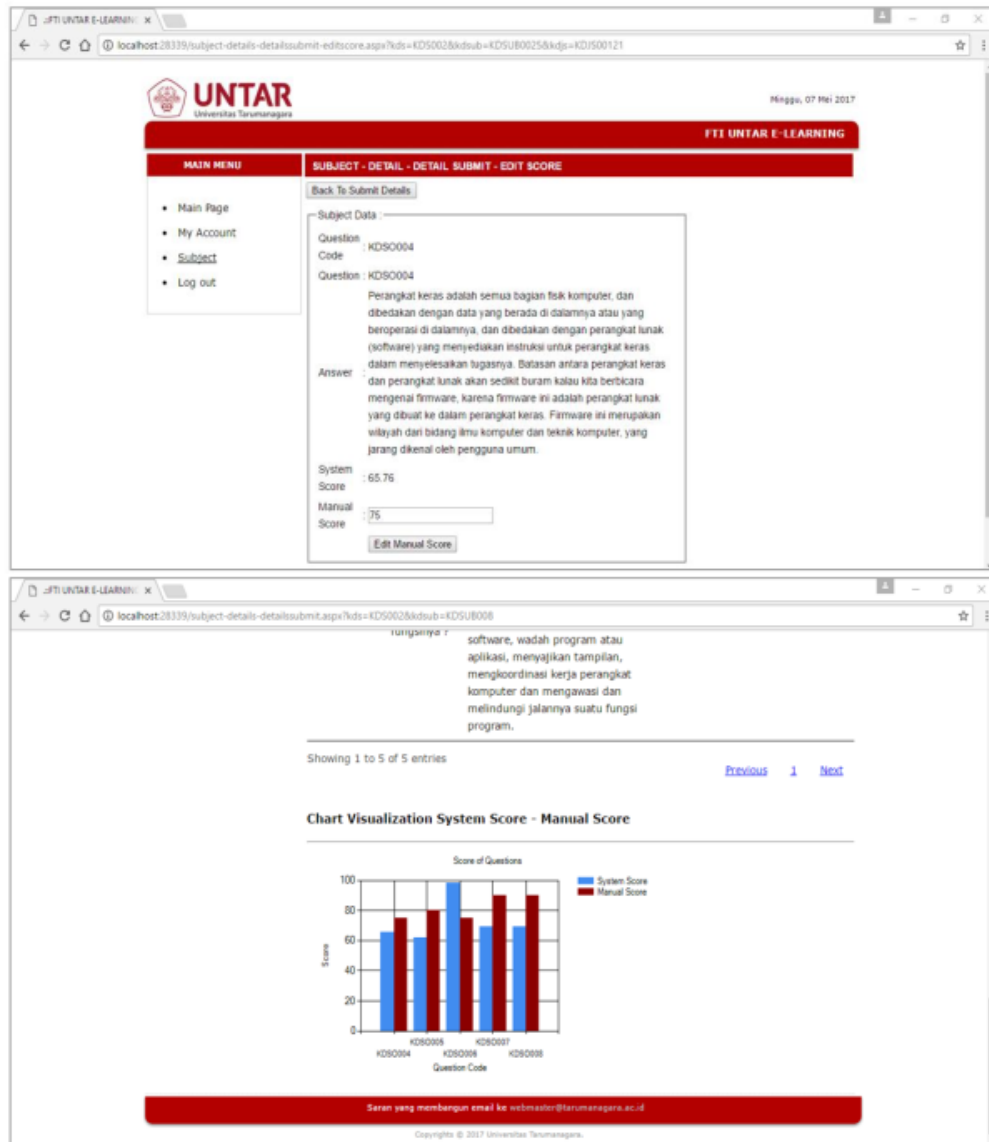


Fig. 4. Testing results of the automatic essay scoring module when the student chooses the exam.



7

Fig. 5. Testing results of the automatic essay scoring module when the student submits the exam.



7

Fig. 6. Testing results of the automatic essay scoring module when the lecturer gives the value manually.

4.2. Automatic essay scoring feature testing

The process of automated essay scoring system is done to find out whether the system created can assess the essay answers according with the assessment conducted by human or lecturers and find out how much the average level of accuracy generated by the automatic essay scoring system. The process of automated essay scoring system test is divided into two parts namely the testing process by the researchers and the testing process by the respondents.

4.2.1 Automatic essay scoring feature testing by researchers

In the process of testing conducted by researcher, the initial process is to collect a set of questions and answers, then the data collected will be tested whether the automatic essay scoring system can assess the essay answers automatically according with the assessment of humans or lecturers. After that the collected data will be validated by Faculty of

Technology Tarumanagara University lecturers. Testing conducted by the researchers there are 2 parts are testing the perfect answer with the same word order and testing the perfect answer with different word order.

4.2.2 Automatic essay scoring feature testing by respondents

In the process of testing by respondents, the testing begins by distributing questionnaires to Faculty of Technology Tarumanagara University students consisting of questions about the course material "Introduction to Information Technology". Total respondents who fill the questionnaire as many as six people while the total answers of students who successfully collected 30 answers. Before the testing started, first the questionnaire will be assessed manually by experts who are lecturers of Faculty of Technology Tarumanagara University. The following are some of the data for questions on the questionnaire:

Table 1. Some of the data for questions on the questionnaire.

| No | Questions |
|----|--|
| 1 | <i>Apakah yang dimaksud dengan perangkat keras?</i> (What is hardware?) |
| 2 | <i>Jelaskan yang anda ketahui tentang internet?</i> (Explain what you know about the internet?) |
| 3 | <i>Jelaskan yang anda ketahui tentang pengertian dan fungsi processor pada komputer?</i> (Explain what you know about the meaning and function of the processor on the computer?) |
| 4 | <i>Jelaskan apa yang dimaksud dengan jaringan komputer beserta tujuannya?</i> (Explain what is meant by computer network and the purpose?) |
| 5 | <i>Apa yang dimaksud dengan sistem operasi pada komputer ? Jelaskan beserta fungsinya?</i> (What is an operating system on a computer? Describe along with the function?) |

Data in the form of questionnaires that have been collected will be tested to find out how much the average mean absolute error between the assessment by the system with the manual assessment by humans and the average mean absolute percentage error of the automatic essay scoring for assessing essay answers. Then with the average level of error percentage can also be measured how much the system accuracy level.

Testing toward the automatic essay scoring feature is conducted through five scenarios, namely:

1. LSA Unigram.
2. LSA Bigram.
3. LSA Trigram.
4. LSA Unigram + Bigram.
5. LSA Unigram + Bigram + Trigram.

Summary results of testing scenarios conducted by respondents can be seen in the Table 2.

Table 2. The results of testing scenarios by respondents.

| Number of Answer | Lecturer Score | System Score | | | | |
|------------------|----------------|--------------|-------------|-------------|-------------|-------------|
| | | Scenario 1. | Scenario 2. | Scenario 3. | Scenario 4. | Scenario 5. |
| 1. | 70 | 63.58 | 43.34 | 24.15 | 62.26 | 55.44 |
| 2. | 90 | 41.14 | 18.78 | 4.49 | 36.34 | 32.68 |
| 3. | 90 | 91.76 | 45.64 | 45.37 | 48.46 | 41.03 |
| 4. | 90 | 62.63 | 20.21 | 13.61 | 41.45 | 36.93 |
| 5. | 80 | 68.35 | 24.88 | 0 | 28.86 | 24.95 |
| .. | ... | ... | ... | ... | ... | ... |
| 30. | 100 | 100 | 52.87 | 47.82 | 91.44 | 86.58 |

4.2.3 Evaluation of automatic essay scoring feature by researchers

Based on the results of testing conducted by researchers, the system has been able to assess the essay answers with the range (1 to 100). Scores that can be generated system is unigram score, bigram score, trigram score, unigram + bigram score, and unigram + bigram + trigram score. Here is one example of testing conducted by researchers.

Questions:

Apa yang dimaksud dengan perangkat keras?

(What is hardware?)

Answer key:

Perangkat keras merupakan bagian fisik dari komputer yang berfungsi untuk memberi masukan, menampilkan dan mengolah keluaran. Perangkat keras ini digunakan oleh sistem untuk menjalankan suatu perintah yang telah diprogramkan artinya berhubungan dengan perangkat lunak. Perangkat keras ini dibedakan menjadi empat jenis yaitu perangkat input, pemrosesan, output dan penyimpanan.

(Hardware is the physical part of the computer that serves to provide input, display and process the output. This hardware is used by the system to run a command that has been programmed which means it is related to the software. This hardware is divided into four types consisting of input devices, processing, output and storage.)

4.2.3.1 Testing the perfect answer with the same word order

Answer:

Perangkat keras merupakan bagian fisik dari komputer yang berfungsi untuk memberi masukan, menampilkan dan mengolah keluaran. Perangkat keras ini digunakan oleh sistem untuk menjalankan suatu perintah yang telah diprogramkan artinya berhubungan dengan perangkat lunak. Perangkat keras ini dibedakan menjadi empat jenis yaitu perangkat input, pemrosesan, output dan penyimpanan.

(Hardware is the physical part of the computer that serves to provide input, display and process the output. This hardware is used by the system to run a command that has been programmed which means it is related to the software. This hardware is divided into four types consisting of input devices, processing, output and storage.)

Unigram score: 100.

Bigram score: 100.

Trigram score: 100.

Unigram + bigram score: 100.

Unigram + bigram + trigram score: 100.

4.2.3.2 Testing the perfect answer with different word order

Answer:

Keras perangkat merupakan bagian fisik dari komputer yang berfungsi untuk masukan memberi, menampilkan dan mengolah keluaran. Keras perangkat ini digunakan oleh sistem untuk menjalankan suatu perintah yang telah diprogramkan artinya berhubungan dengan lunak perangkat. Keras perangkat ini dibedakan menjadi empat jenis yaitu perangkat input, pemrosesan, output dan penyimpanan.

(Device hard is the physical part of the computer that serves to provide input, display and process the output. This hardware is used by the system to run a command that has been programmed which means it is related to the software. This hardware is divided into four types consisting of input devices, processing, output and storage.)

Unigram score: 100.

Bigram score: 37.04.

Trigram score: 33.90.

Unigram + bigram score: 83.38.

Unigram + bigram + trigram score: 72.91.

4.2.4 Evaluation of automatic essay scoring feature by respondents

Based on the results of testing conducted by respondents, the testing results for the automatic essay scoring system by the respondents can be seen in Table 3 and the level of system accuracy in the test scenario is illustrated in Figure 7.

Table 3. The results of the automatic essay scoring system by respondents.

| No | Testing Scenarios LSA | MAE | MAPE |
|----|--------------------------------|----------|---------|
| 1. | LSA Unigram | 16.67267 | 21.35 % |
| 2. | LSA Bigram | 45.06733 | 58.89 % |
| 3. | LSA Trigram | 63.71067 | 85.09 % |
| 4. | LSA Unigram + Bigram | 22.42033 | 28.63 % |
| 5. | LSA Unigram + Bigram + Trigram | 27.38000 | 35.21 % |

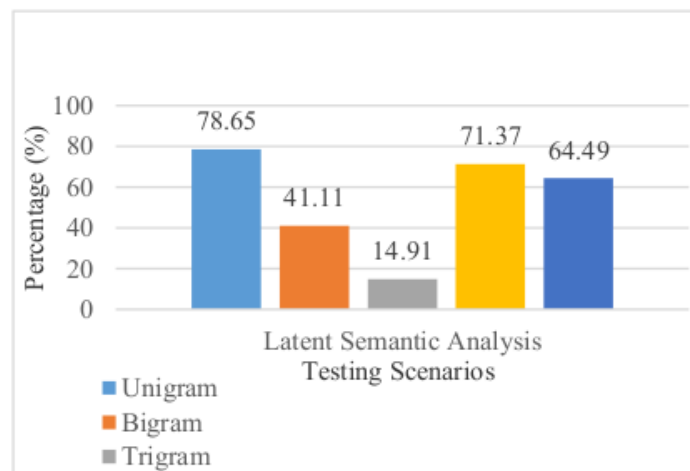


Fig. 7. Graph the average rate of system accuracy percentage by respondents.

Based on the test results the average accuracy rate of the resulting assessment of automatic essay scoring system is quite good or almost similar for value provided by the lecturer. In addition, from five testing scenarios performed, the test with the unigram term gives the best results with an accuracy of 78.65 % or a system error rate of 21.35 % and the mean value difference between the manual value and the system value is 16.672. In the testing process with bigram, trigram, unigram + bigram, unigram + bigram + trigram the resulting accuracy is 58.89 %, 14.91 %, 71.37 %, 64.79 % with average value difference of 45.067, 63.710, 22.420, 27.380. The cause of small accuracy in the testing process of bigram terms and trigram term testing is the rarity of the frequency of occurrence of word or term bigram or trigram found in the lecturer answer key. Another thing that affect the accuracy level of the automatic essay scoring system is:

- i) There are still many error words in the answer when the system created can not handle the error correction word.
- ii) Besides the other cause is the lack of an alternative answer key that is incorporated into the system.
- iii) In addition, from the answers students still use a lot of words abbreviated for example "Internet Protocol" abbreviated to IP.
- iv) Another thing that causes the difference of judgment is that the answers key entered by the system can have a similarity of meaning or understanding with other words such as there are similarities of words or words in a foreign language.
- v) Another cause is the assessment of the system is very affected on the number of words that exist in each sentence. The results of the system assessment are 11 average value of similarity between term relationship with the sentence between the student's answer with the lecturer answer key. So that the assessment system will see the completeness of an answer based on each relationship on a term with a sentence whereas in the manual assessment if there is one sentence correct answer then the value given high. This is because the manual assessment sees the completeness of the answer overall than the whole sentence.

17

5 Conclusion

Based on the results of tests that have been done, can be obtained conclusions on the system that has been made as follows:

- i) Each module on the website has been running well.

- ii) The automatic essay scoring system has been able to assess student essay answers.
- iii) Assessment of the essay answers generated by the automatic essay scoring system is influenced by the presence or absence of key term answers in a sentence, if the more the same term distribution of keywords between answers with answer key then the resulting value will be high, and vice versa.

References

1. S. Deerwester, S.T. Dumais, R. Harshman, *Indexing by latent semantic analysis* [Online] from <http://lsa.colorado.edu/papers/JASIS.lsi.90.pdf> (1992) [Accessed on 17 September 2017]
2. S. Dikli. *JTLA*, **5**,1 (2006). <https://files.eric.ed.gov/fulltext/EJ843855.pdf>
3. M.M. Islam, A.S.M. Latiful Hoque. *JCP*, **7**,3:616–626 (2012). <http://www.jcomputers.us/index.php?m=content&c=index&a=show&catid=125&id=1969>
4. R.H Gusmita, R. Manurung, *Penerapan latent semantic analysis untuk menentukan kesamaan makna antara kata dalam Bahasa Inggris dan Bahasa Indonesia [Applied latent semantic analysis to determine same meaning between words in English and Indonesian]* [Online] from http://digilib.mercubuana.ac.id/manager/t!@file_artikel_abstrak/Isi_Artikel_39179235_2024.pdf. (2008) [Accessed on 25 January 2018]. [in Bahasa Indonesia]
5. Ch.A. Kumar, S. Srinivas, *CIT*, **17**,3:259–264 (2009). <http://cit.fer.hr/index.php/CIT/article/view/1751>
6. R.B.A. Pradana, Z.K.A. Baizal, Y. Firdaus. *Automatic essay grading system menggunakan metode latent semantic analysis*, Seminar Nasional Aplikasi Teknologi Informasi (Yogyakarta, Indonesia 2011). Seminar Nasional Aplikasi Teknologi Informasi (SNATI):E78-86 (2011). <http://journal.uii.ac.id/index.php/Snati/article/view/2217>
7. S.A. Sugianto, Liliana, S. Rostianingsih. *Jurnal Infra*, **1**,2:119–124 (2013). <https://www.neliti.com/id/publications/105718/pembuatan-aplikasi-predictive-text-menggunakan-metode-n-gram-based>
8. Munir, L.S. Riza, A. Mulyadi. *IJTRD*, **3**,6:403–407 (2016). <http://www.ijtrd.com/papers/IJTRD5412.pdf>
9. N.T. Thomas, A. Kumar, K. Bijlani. *Procedia Computer Science*, **58**:257–264 (2015). https://ac.els-cdn.com/S1877050915021304/1-s2.0-S1877050915021304-main.pdf?_tid=29855a6a-fed1-11e7-9742-00000aab0f6b&acdnat=1516556184_8c82910250d80802478e13aedd6c965f
10. T. Kakkonen, N. Myller, E. Sutinen. *Applying part-of-speech enhanced LSA to automate essay grading*, Proceedings of the 4th IEEE International Conference on Information Technology: Research and Education (Tel Aviv, Israel, 2006). <https://arxiv.org/abs/cs/0610118>

18%

SIMILARITY INDEX

13%

INTERNET SOURCES

11%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

| | | |
|---|--|-----|
| 1 | www.icesti.org Internet Source | 4% |
| 2 | Submitted to Universitas Brawijaya Student Paper | 4% |
| 3 | markastekno.blogspot.com Internet Source | 2% |
| 4 | Viny Christanti Mawardi, Niko Susanto, Dali Santun Naga. "Spelling Correction for Text Documents in Bahasa Indonesia Using Finite State Automata and Levinshtein Distance Method", MATEC Web of Conferences, 2018 Publication | 2% |
| 5 | Submitted to Universitas Pendidikan Indonesia Student Paper | 1% |
| 6 | www.e3s-conferences.org Internet Source | 1% |
| 7 | Errissya Rasywir, Yovi Pratama, Hendrawan, Marrylinteri Istoningtyas. "Removal of Modulo as Hashing Modification Process in Essay Scoring System Using Rabin-Karp", 2018 International Conference on Electrical Engineering and Computer Science (ICECOS), 2018 Publication | <1% |
| 8 | Julianto Lemantara, M. J. Dewiyani Sunarto, Bambang Hariadi, Tri Sagirani, Tan Amelia. "Prototype of Automatic Essay Assessment and Plagiarism Detection on Mobile Learning | <1% |

"Molearn" Application Using GLSA Method",
2019 International Seminar on Research of
Information Technology and Intelligent Systems
(ISRITI), 2019

Publication

-
- | | | |
|---|--|-----|
| 9 | docplayer.net Internet Source | <1% |
|---|--|-----|
-
- | | | |
|----|--|-----|
| 10 | Islam, Md. Monjurul, and A. S. M. Latiful Hoque. "Automated Essay Scoring Using Generalized Latent Semantic Analysis", Journal of Computers, 2012. Publication | <1% |
|----|--|-----|
-
- | | | |
|----|--|-----|
| 11 | A Amalia, D Gunawan, Y Fithri, I Aulia. "Automated Bahasa Indonesia essay evaluation with latent semantic analysis", Journal of Physics: Conference Series, 2019 Publication | <1% |
|----|--|-----|
-
- | | | |
|----|---|-----|
| 12 | Submitted to Universitas 17 Agustus 1945 Surabaya Student Paper | <1% |
|----|---|-----|
-
- | | | |
|----|---|-----|
| 13 | Submitted to Kingston University Student Paper | <1% |
|----|---|-----|
-
- | | | |
|----|--|-----|
| 14 | discovery.ucl.ac.uk Internet Source | <1% |
|----|--|-----|
-
- | | | |
|----|--|-----|
| 15 | Submitted to University of Pune Student Paper | <1% |
|----|--|-----|
-
- | | | |
|----|--|-----|
| 16 | academicjournals.org Internet Source | <1% |
|----|--|-----|
-
- | | | |
|----|---|-----|
| 17 | Anton Yudhana, Sunardi -, Abdul Djalil. "Implementation of Pattern Matching Algorithm for Portable Document Format", International Journal of Advanced Computer Science and Applications, 2017 Publication | <1% |
|----|---|-----|
-

18

www.math.uky.edu

Internet Source

<1%

19

"Mining Intelligence and Knowledge Exploration", Springer Science and Business Media LLC, 2013

Publication

<1%

20

archive.org

Internet Source

<1%

21

awidiyar.smplematanglestari.sch.id

Internet Source

<1%

22

www.stat.cmu.edu

Internet Source

<1%

23

Kevin R. Parker, Robert Francis Williams, Philip S. Nitse, Siong Meng Albert Tay. "Use of the Normalized Word Vector Approach in Document Classification for an LKMC", Issues in Informing Science and Information Technology, 2008

Publication

<1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On