A Combination of Text Similarity And Latent Semantic Analysis (LSA) Methods In Automatic Scoring Essay

Deddy Atmajaya, Purnawansyah, Siti Rahayu

ABSTRACT--- Description of type exam (essay) is considered by many experts as the most appropriate test to reap the results of a complex learning activities, because essay writing will involve the student's ability to remember, organize, express, and integrate the ideas of the students. Just to correct the essay exam results, requiring a longer time if done manually because most do by reading an essay one by one. So that, lecturers needs to spend a lot of time to assess the answers of student's exam. Therefore, in implementation, automatic scoring system is needed on the answer essay exam. Automated essay assessment method used in this study is a combination of Text Similarity and Latent Semantic Analysis (LSA) to look for a match and similarity level student answers with the answer key that has previously been inputted into the system. Data used in this test is 15 students with each student to answer 5 questions. Data obtained from subjects essay Basis Data I. The correlation results of that two assessment shows grades 0,946085 with an average increment of 2,08. Which means the results of the assessment system is not much different from the results of the assessment of the lectures, so that the automatic scoring system can be applied to essay type exam.

Kewords: Automatic scoring essay; Text similarity; Latent semantic analysis.

I. INTRODUCTION

In the era of information technology developed at this time, the test would have been able to be carried out easily with the use of technology as a medium that is one of them done online. Trials with various types (multiple choice, essay, short stuffing) are applicable to online exams [1]. But they rarely are online examination system which applies essay exam with a correction carried out directly by the system. Unlike the exam multiple choice or multiple choices that have often directly can be corrected by the system because it is objective and value of the test results can be directly obtained, essay exams need to go through the correction process by the lecturers / maker about to give the value manually inputted into the system, then the system displays the interface in charge of examinees.

Automated Essay Assessment is the process of scoring on essay tests performed by the system so that the value of the essay exam results can be directly obtained in real time. On this case the lecturers only need to input questions, answer key and the weight of the question, and the system then served to correct exam results and final scores based on the calculation of weights that have been set by the presentation of the degree of similarity is entered student answers with the answer key lecturers. This makes the process of correcting the results of essay exams easier and more practical.

The combination of Text Similarity and Latent Semantic Analysis (LSA) methods can solve this problem [2]. On the use of text similarity method text will answer a series of processes to find keywords and two text answers are tested will be matched using Latent Semantic Analysis (LSA) [3], [4].

This research will be carried out the application design by comparing the similarity essay with an answer key that already existed in the system. By knowing the percentage of similarity of the two texts of these answers can be used as reference for the assessment. Then it will do a comparison between the responses of the system (rater by system) with manual assessment (human rater).

II. EXPERIMENTAL DETAILS

Pre-processing stage is the stage where the selection of application data to be processed on each document. This pre-processing process includes folding case, tokenizing, filtering, and stemming. Figure 1, is a phase of pre-processing.

TF-IDF is one of the weight calculation of the frequency of occurrence of a term in the document. TF-IDF is used to find the value of the weight of the document. TF-IDF algorithm will examine the appearance of each word on the content of a document from the tokenizes, filtering, and stemming from each word in the document's contents [5].

Retrieval Number:E12060585C19/2019©BEIESP
DOI: 10.35940/ijeat.E1206.0585C19

1432

Published By:
Blue Eyes Intelligence Engineering
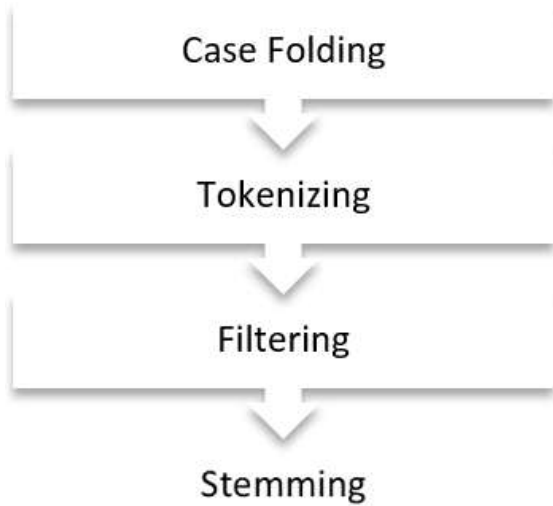*& Sciences Publication*

Fig. 1: Text Pre-processing Flow

The method is implemented in Automated Essay assessment process is a combination of Text similarity and Latent Semantic Analysis methods which outlines the process flow is shown in the Figure [6].
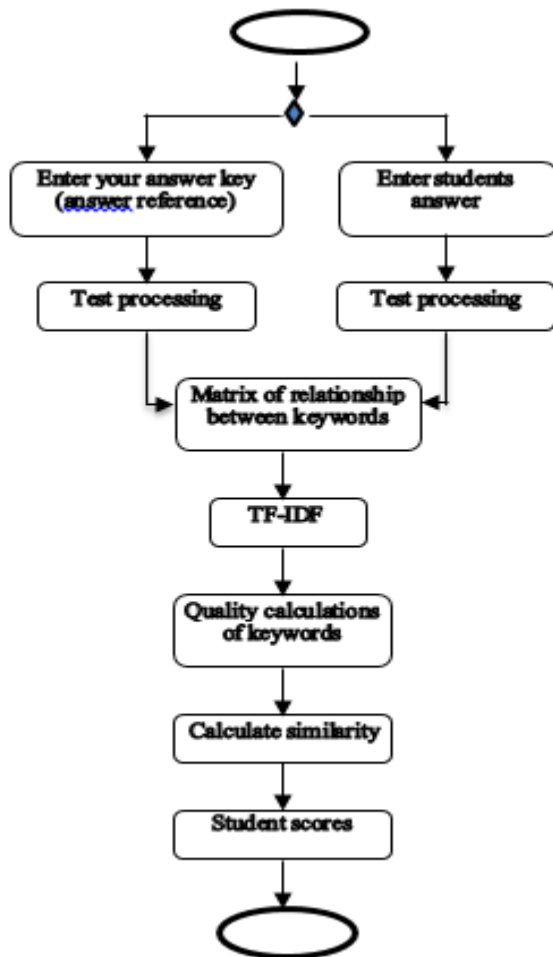

Fig. 2: Flow of Value Calculation Process

Based on the above diagram can be seen that the value obtained by the students is the result of matching the relationship between keywords and then calculated the weight based on the TF-IDF matching results and the final step is to determine the level of accuracy using the cosine similarity and similarity answer. An important component in the assessment essay on this system is the answer to the reference, the student answers and processes determine the relationship between keyword matrix reference answers and answers students.

III. RESULT AND DISCUSSION

To find out how good the quality of the assessment carried out by the system, it will be done the comparison results of the assessment performed by the system with the results of assessment of lecturers. Performance assessment will be based on the magnitude of the correlation and the average difference between the systems with the lecturer [7].

Table 1: Results of the assessment (lecturers) and appraisal system to some students (personal)

Name	Class	Q1		Q2		Q3		Q4		Q5	
		S	D	S	D	S	D	S	D	S	D
Citra Suardi	B2	20	20	20	25	16	20	13	15	15	15
Issahtul A. A	B2	18	20	23	25	22	25	15	15	14	5
Halmiah	B2	20	20	23	15	21	25	14	15	12	10
Herawati S.	B2	19	20	22	25	21	25	15	15	13	5
Indah S. K.	B2	20	20	23	25	24	25	15	15	12	5

Table 2: The average value of the assessment results lecturers and systems

The Average Result of thr Assessment System (a)	The Average human rater assessment (b)	Difference (a) and (b)
19.4	20	0.6
22.2	23	0.8
20.8	24	3.2
14.4	15	0.6
13.2	8	5.2

Based on the results of data processing, it can be compared to the correlation and the percentage difference from the results of the assessment system and human raters. From the data obtained the correlation of the two ratings 0.94609 figures show the average difference between the value of 2.08. By looking at the correlation value Table 3.

Table 3: Correlation coefficients and interpretation

Sample Correlation Values	Interpretation Values
0.00-0.09	Negligible correlation relationship
0.10-0.29	Low correlation relationship
0.30-0.49	Moderate correlation relationship
0.50-0.70	Medium correlation relationship
>0.70	Strong correlation relationship

So it can be said that the relationship is very strong correlation and response system assessment results correlate with the results of assessment of lecturers manually. It means that the assessment system is not much different from the results of the assessment of lecturers, so that the automatic scoring system can be applied to essay-type exam.

IV.   CONCLUSION

Based on the result, it can be concluded that an assessment answers produced by the system is influenced by several factors. The factors are (a) presence or absence of key words in the sentence answers to students. (b) The more differences in the number of words in a sentence between the assessments made by the lecturers to students, then the resulting value is smaller. (c) If a student has a spread-sentence answer increasingly keyword is the same asthe answer key lecturers, the resulting value is also higher. (d) The system is not able to detect any intent or synonyms same answer between students and professors.

The use of the Text Similarity and Latent Semantic Analysis methods in the essay assessment resulted in a correlation between the assessments made by the system with a high lecturer appraisal, indicated by the number 0.946085 with the average value difference for each question of 2.08. This means that the results of the system assessment do not vary much with the results of the lecturer's assessment, so that the implementation of the combination of Text Similarity and LSA methods can be said to have been able to provide good results on the assessment of essay type tests.

In this study, the need for additional processes to detect the order of words in a sentence that assessment is not only based on keywords alone, but the order of words in a sentence is also worth noting, keep the meaning of the sentence unchanged. Furthermore, it should be added the correction system in the form the mathematical answers / calculation, using special symbols, etc.

V.   ACKNOWLEDGEMENT

REFERENCES

1. D. Metzler, S. Dumais, and C. Meek, "Similarity measures for short segments of text," in European conference on information retrieval, 2007, pp. 16–27.
2. A. M. Yehia, I. L. F., and M. F. Abulkhair, "Text mining and knowledge discovery from big data: challenges and promise," Int. J. Comput. Sci. Issues, vol. 13, no. 3, p. 54, 2016.
3. K. P. Kumar, V. N. Mandhala, S. Vempati, and S. R. Peram, "Finding author similarity by clustering probabilistic LSA factors in INDIAN english authors poetry," Int. J. Eng. Technol., vol. 7, no. 2.7, pp. 1096–1099, 2018.
4. S. S. Kamaruddin, Y. Yusof, N. A. A. Bakar, M. A. Tayie, and G. A. A. J. Alkubaisi, "Graph-based Representation for Sentence Similarity Measure: A Comparative Analysis," Int. J. Eng. Technol., vol. 7, no. 2.14, pp. 32–35, 2018.
5. G. Jorge-Botana, J. A. Leon, R. Olmos, and I. Escudero, "Latent semantic analysis parameters for essay evaluation using small-scale corpora," J. Quant. Linguist., vol. 17, no. 1, pp. 1–29, 2010.
6. T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," Discourse Process., vol. 25, no. 2–3, pp. 259–284, 1998.
7. T. J. Miller et al., "Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability," Schizophr. Bull., vol. 29, no. 4, pp. 703–715, 2003.