

The Valuation of Otomatic Essay Answer by Similarity Jaccard Method at Al Firdaus Vocational High School

Muhammad Djaenudin¹, Heru Purwanto², Rahayu Widiono³, Mumtaz Muttakin⁴, Rizky Miftahul Fajri⁵.

¹Master of Computer Science Study Program, Budi Luhur University

²Master of Computer Science Study Program, Budi Luhur University

³Master of Computer Science Study Program, Budi Luhur University

⁴Master of Computer Science Study Program, Budi Luhur University

⁵Master of Computer Science Study Program, Budi Luhur University

Email: ¹whyjay3@gmail.com, ²h3ru.purwant0@gmail.com, ³andreaswwt@gmail.com,
⁴mumtazmuttakin@gmail.com, ⁵kikihootowl@gmail.com

ABSTRACT

Testing with essay system is form of evaluation while answer choice is not available, and student must answer with the sentence, so as can exercise student in presenting an information verbally, beside the essay test also demands better comprehension about knowledge and can be used to measure level of man comprehension about the science deeply. Testing with this system becomes selection of teacher to evaluate the level of student compherension ability even though in reality is not easy to give objective assessment on student's answer. Teacher needs a lot of time to check essay's answer, more the number of testing and more the number of student follows the test, so more the number of testing will be checked by teacher. This cause the quality of assessment decreasing and sometimes the assessment is not objective anymore. Various of method that can be used to carry out otomatic essay answer assessment one of them is using Jaccard Similarity Method. Jaccard Similarity Method is a method to calculate similarity or identical from two documents, in this research is used student testing answer document and teacher answer key document. The result of this research is found the average value from expert is 88.91 and average value from the system is 85.52, the different average value from expert and system namely $88.91 - 85.52 = 3.39$. The accuracy which is gotten from Jaccard Similarity method in this research is 96.19%.

Keywords ~~Essay~~ **Test, Jaccard similarity, Text similarity, Word Quality**

1. INTRODUCTION

Assessment in education is a process to measure the achievement of learning and teaching process. Assessment is an important thing in learning process. Assessment usually be carried out based on three main aspects, namely skill, knowledge, and attitude. Assessment consists of test and non-test [1]. The assessment included in non test namely observation, interview, and attitude scale. The test assessment is devided into written test and oral test.

Written test in general is classified into objective and subjective tests. Subjective test consists of short argument test and long

argument test. Written test can be given in form of multiple choice questions and essay. Written test in form of essay is choosen because it can more measure how far the students comprehend and able to explain about the materials have been gotten.

Evaluation result is an indicator from student's comprehension about the materials taught. Learning evaluation type can be a multiple choice that has prepared some choice answer while the students only choose the most exact answer. The type of other evaluation namely the essay evaluation type with no choice answer given. Generally the evaluation in form

of essay needs the answer such as discussion, argument, or explanation.

The forms of objective nor subjective tests (essay) is arranged in order to measure trully student ability. About essay, a teacher will give written's question then the students is asked to answer the question in form of description writting in their own way. Essay test has advantage in measuring cognitive aspect, to develop language ability, to train regular thinking or reasoning and develop the skill of problem solving [2].

Different from written test of multiple choice, the essay written test will need extra time to carry out correction or answer assessment. A teacher has to read one by one of answer pattern from the student to then be accordanced with available key answer. Objectivity and consistency of teacher in giving the value is very important while must decide a right or wrong answers because it will influence against assessment results. The worst of essay question is long time of correction and cannot be represented by others. [3].

Another problem will be faced by teacher is hard to give objective value against essai answer. So that, it's needed a solution which able to ease implementation of essay anser assessment by not ingnore objectivity assessment. So researcher argues that development about subjective test question needs to be researched further.

Research related to otomatic essai answer assessment was ever carried out before with subject Otomatic Essay Answer Assessment System Based On Sentence Closeness Value, output from this assessment system is a closeness grade or value between answer and key answer. An answer will have high grade if closeness value between answer with answer key is high, otherwise if closeness value between answer and answer key is low, so the grade resulted is also low. The kind of semantic used in this research is semantic relatedness based on dictionary or known as Mechine Readable Dictionary (MRD). By using MRD System will find semantic relation among words based on dictionary. The step of first otomatic assessment system testing is to count total value (grade) of each student. Comparing total value

resulted from the system with given value by instructure and counting how the accuration. From comparison result of assessment given by instructure against system is gotten the accuration value is 83%. [9].

Based on background and problem faced at Al Firdaus Vocational School, so it need to be carried out research by build otomatic essay answer assessment system to ease teacher in doing assessment against the test in form of essay answer. The system built of the testing is still limited on the lesson of Simulation and Digital Communication. By building the system, is expected can introduce the essay test based on computer as testing media and as answer assessment to accelerate learning evaluation process.

2. THEORY BASE LEARNING RESULT ASSESSMENT

Assessment is a process giving or determining value to certain object based on a certain criteria, the process of giving on going value in form of interpretation that ended by judgement[2]. Assessment is conversion from raw score in to evaluative classification according to relevant norm or criteria [4]. Evaluation classification based on a class norm or criteria is determined previously. Assessment against learning result can be classified into two approach namely refer to the criteria and assessment refer to norm.

Assessment function is: equipment to know the accomplished or not the instructional aim, feedback for improvement learning teaching process amd base in arranging student learning advance report for student's parents [2]. Assessment function performance observation of activity components of learning teaching process in achieving expecting target in learning teaching process [5]. Assessment function can be concluded to observe and to know the development of the component in learning and teaching activity so as can be taken further act accordance with objective determined.

Test assessment consist of objective and subjective tests. Test is a question to get answer from student in form of oral, written or action [2]. The field measured uses generally test is

cognitive or acknowledge fields. Origin test also can to measure affective and psychomotoric fields.

Objective test is most testing used to value learning result. Objective test is chosen because of widen lesson material scope and easy in doing assessment. The form of objective questions to wit:

- 1) short answer,
- 2) true – false,
- 3) matching,
- 4) multiple choice [2].

Test form beside short answer in question has available the possibility answer can be chosen.

Essay test is one of the oldest test model and commonly used in learning evaluation. Essay test is question that assert the student's answer in form of description, explanation, discussion, comparison, giving reason, and other same forms with assertion question by using own's word and language [2].

Essay test model is divided into:

- 1) free essay test (free description)
- 2) limited and structured essay test [2].

The student's answer in free essay test is not limited. Limited essay test has question form which is steered on certain things or certain available limitation. The others essay test model is structured essay test, this test is shown as form between objective question and essay question. Structured test is short answer question, even the answer is free but the question has elements: a) question introduction, b) set of data, c) series of sub-question.

The advantage and the shortcoming of essay test model are: a) can measure high mental process or high level of cognitive aspect; b) can develop language capability, both oral or written in good manner accordance with language rules; c) can exercise thinking ability in neat or reasoning, namely logic thinking, analitic and systematic; d) developing the problem solving capability; e) There is a technical advantage such as an easy in arranging the question, teacher can directly sees the students thinking process. The shortcoming of essay test question to wit: a) Test sample is very limited so as it would be impossible to test all of materials; b) It is very subjective, both in questioning, making question or checking it; c)

The test is lack reliable, to disclose limited aspect, the checking needs lenght of time.

TESTING BASED ON COMPUTER

Assessment using computer is used to replace PPT (paper-pencil test). Computer utilizing to do the test is called with Computerized Based Testing (CBT), namely in principal is same with PPT but the question is given with computer assistance. The serving with computer assistance is possibly carried out randomization of given question, so the opportunity of test's participant to cheat becomes less.

Testing based on computer is divided into several types, to wit: computerized based testing (CBT), fixed form computerized based testing, linear on the fly test (LOFT), computerized adaptive testing (CAT) [6]. Place in securing item of question in CBT is divided into 1) in individual computer, 2) in server computer through local area network (LAN), 3) in server computer through internet.

Computerized based testing is a simplest model from testing based on computer. Testing is (CBT) managed and generally the assessment is carried out electronically. Various types of questions can be shown through computer's monitor and is conversion from test by using paper media. Question model such as objective or subjective questions. The question may be saved in individual computer, local area network (LAN), or in outside server computer with internet communication media. CBT makes possibility the test can be managed and be valued more accurate, faster, and more secure from the test using the paper.

Fixed form CBT is a test while all of test's participants get the same questions. Fixed form CBT is model of paper and pencil test (PPT) which is in serving is converted using computer assistance. The pattern in serving fixed form CBT test which differentiate it with PPT is by conversion of ordinal in serving item of questions to the test's participants. The opportunit in cheating in model fixed form CBT is big enough because the item of question is same in content. The advantage of fixed form CBT of item of question which is developed relatively is more less.

LOFT is a further development from CBT. On LOFT's computer has item of questions collection in a lot number and has variation according to level of difficulty of item of questions. Test's participants on LOFT's model get variation of question's item and it's different for each test's participants. Question's item of LOFT's model which is served even different in content but in characteristic of statistic have the same quality. The advantage of this test namely the test's participant has little chance to cheat because the different of question's item. The lack of LOFT namely the item of question which is developed for question's bank is lot in number.

Essay's test with computer's assistance is form of CBT which is developed by researcher. tes Essay's test uses computer's assistance to serve the test and value it. Essay's test with computer's assistance has two main functions namley to make essay question and to value the answer of essay's test. The essay's answer in carrying out the correction uses string similarity method namely by comparing the question's answer with existing answer in question's bank.

TEXT MINING

Text mining is a extraction process of pattern (useful information and knowledge) from big enough of unstructured data source. Text mining has target and using the same process with data mining, but has different input. Input for text mining is a less or unstructured data, such as Word document, PDF, text quote, etc. While the input for data mining is unstructured data[7]. The main aim of text minig is to support the process of knowledge discovery in big document collection.

In principal, text mining is a multidiscipliner science field, including information retrieval (IR), text analysis, information extraction (IE), clustering, categorization, visualization, database technology, natural language processing (NLP), machinelearning, and data mining. There are four main phases in text mining, namely text preprocessing, text transformation, feature selection, and pattern discovery.

Case folding is a phase that converts all letters in document become little letters. Only letter from 'a' till 'z' being accepted. Character beside letter is erased and considered as delimiter [8]. Example of the using of case folding, if the sentence that is installed as "Norm of honest behavior, responsible, and aware related with utilization of Information Technology and Communication together " so the output will become "norm of honest behavior, responsible, and aware related with utilization of information technology and communication together", for further description can be shown on table 1.

Table 1: Case Folding Process

	Original Text	After Case Folding Process
Answer Key	<i>Norm of honest behavior, responsible, and aware related with utilization of Information Technology and Communication together</i>	<i>norm of honest behavior, responsible, and aware related with utilization of information technology and communication together</i>
Student Answer	<i>Norm of honest behavior, responsible, and aware related with utilization of information technology and communication together</i>	<i>norm of honest behavior, responsible, and aware related with utilization of information technology and communication together</i>

Table 2: The Result of Tokenizing Process

Answer Key	Student Answer
norm	norm
behavior	behavior
honest	honest
be responsible	and
answer	responsible
and	answer
aware	that
related	be implemented
with	in
utilizing	utilizing
information	technology
and	information
technology	and
communication	communication
in	together
together	same

Table3: Stopword Removal Process and Stemming Output

Answer Key	Student Answer
norm	norm
behavior	behavior
honest	honest
aware	implement
relate	utilize
utilize	technology
information	Information
technology	communication
communication	same
same	

Tokenizing phase is a string input cutting phase based on each letter that arrange it. The process result from this phase can be shown on table2.

Stopword removal yaitu namely throwing the less important letters. And stemming namely carrying out others letters grouping that has basic letter and identical meaning but has different form because of getting different prefix. For further description the result of stopword removal and stemming process can be shown on table 3.

Jaccard similarity has function to search the same letter comparison among two sentences. Jaccard is number of same letters among answer key and student's answer rather than reverse with total letters number as can be shown on formula 1 that can be simplified such on formula 2. Such as cosine distance and matching coefficient, generally this method calculation is based on vector space similarity measure.

$$J(X, Y) = \frac{\sum_{i=1}^p x_i y_i}{\sum_{j=1}^p x_j^2 + \sum_{j=1}^p y_j^2 - \sum_{i=1}^p x_i y_i}$$

$$J(X, Y) = \frac{|x \cap y|}{|x \cup y|}$$

While x is an answer key and y is student's answer. As known on table 3 that are 9 the same letters number from the two answers, and 3 not same and totally are 12 letters.

So $Similarity(X, Y) = \frac{9}{12} = 0,75$

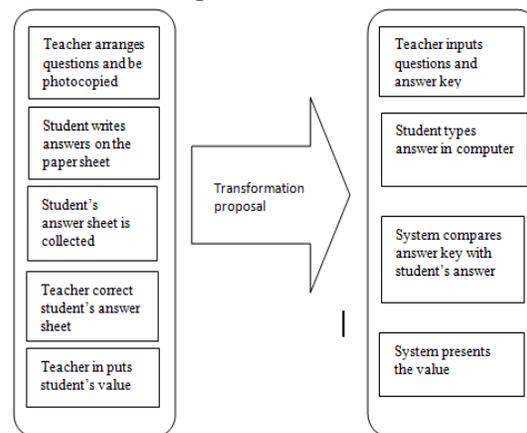
Final value is gotten from similarity value based on formula 2 multiplied with quality of question,

$$Similarity(X, Y) \times \text{quality of question} = 0,75 \times 100 = 75$$

So, the value of students answer against answer key with similarity jaccard method is gotten that the value is 75.

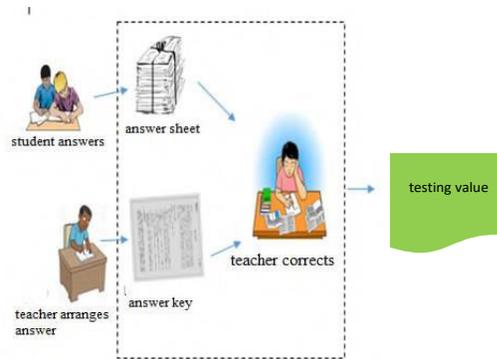
3. DESIGN SYSTEM AND APPLICATION

The process modeling is used to give picture of business process based on user's point of view. Process modeling can be shown from essay test system based on paper at this time and essay test system based on computer that's proposed to ease teacher in carrying out assessment. Essay test process that's proposed can be shown on picture 1.



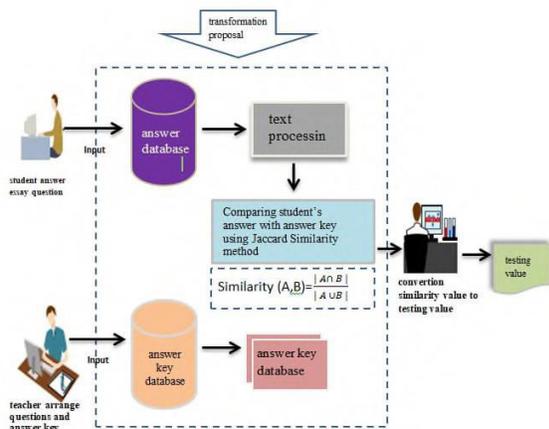
Picture 1: The proposed Essay Test Process

The proposed business process picture uses Otomatic Essay Answer Assessment System which is described that the running business process at the moment such as shown on picture 2. Student answers the essay question has been prepared by teacher, then answer sheets from all students are collected, meanwhile teacher arranges answer and prepare it to then proofread each item of answers on students anwer sheets had been collected one by one, and then can be determined value from each answer sheet.



Picture 2: The running business process at the moment

Then for proposed business process can be shown on picture 3. Student answer available essay question in computer by typing answer in the prepared form accordance to the number. After finish then the answer from each student will be saved in answer database. While teacher arrange question and answer key that will be tested and save it into the database in computer further will be used as server in the essay test implementation. After finish of test implementation while students had answered all available questions in the essay test so the system will carry out its duty, first by preprocessing, and then search similarity value by using jaccard similarity method.



Picture 3: The proposed business process of otomatic essay test

4. OUTPUT AND DISCUSSION

System testing implementation process is carried out in classroom of Al Firdaus Vocational Hight School Jakarta. AK Followed by 25 participants as respondent, 5 of them are Teacher, 3 permanent teachers, and 2 non-

permanent teachers, and 20 students of Class XII accounting. Researcher carries out presentation and demo of otomatic essay answer assessment system which is being developed and explains each available function in the system. After knowing the way in operating this otomatic essay answer assessment system, then respondent is given opportunity to try directly the application. Then respondent is asked to give respon and assessment by filling available form. In this testing process thoroughly asks respondent to give respon against hypotesis in research, that shows whether all of system functions have already accordanced with the need.

The result of otomatic essay assessment system with jaccard similarty method had been carried out against 20 students and their final values such as been shown on table 4 gets average value of 85.52.

Then value that've been carried out by expert which is represented by teacher such been shown on table 5 while teacher carries correction by reading student's answer one by one, compare it based on answer's key, then gives the objective value.

Table 4: Jaccard similarity value Result

No	Respondent	Jaccard Value					final value
		Q1	Q2	Q3	Q4	Q5	
1	Respondent 1	0,75	0,80	0,78	0,75	1,00	78,22
2	Respondent 2	0,75	1,00	0,90	0,91	1,00	91,18
3	Respondent 3	0,67	1,00	1,00	0,82	0,50	73,03
4	Respondent 4	0,33	0,80	0,78	0,91	0,80	72,40
5	Respondent 5	0,50	1,00	0,89	0,92	0,57	75,57
6	Respondent 6	0,88	1,00	1,00	0,75	0,80	88,50
7	Respondent 7	0,63	1,00	1,00	1,00	0,83	89,17
8	Respondent 8	0,75	0,80	1,00	1,00	1,00	91,00
9	Respondent 9	0,75	1,00	0,89	0,91	1,00	90,96
10	Respondent 10	0,75	1,00	1,00	1,00	1,00	95,00
11	Respondent 11	0,67	1,00	1,00	0,91	1,00	91,52
12	Respondent 12	0,50	1,00	1,00	1,00	1,00	90,00
13	Respondent 13	0,88	1,00	1,00	0,91	1,00	95,68
14	Respondent 14	0,75	1,00	1,00	0,75	0,67	83,33
15	Respondent 15	0,67	0,67	0,90	1,00	1,00	84,67
16	Respondent 16	0,33	0,80	1,00	1,00	0,83	79,33
17	Respondent 17	0,75	1,00	1,00	0,75	1,00	90,00
18	Respondent 18	0,75	1,00	0,89	1,00	0,50	82,78
19	Respondent 19	0,88	1,00	0,89	0,75	0,80	86,28
20	Respondent 20	0,63	0,67	0,80	1,00	1,00	81,83
Average Value							85,52

Table 5: Expert (teacher) valuation Result

No	Respondent	Expert Value					Final Value
		Q1	Q2	Q3	Q4	Q5	
1	Respondent 1	80	90	75	80	100	85,00
2	Respondent 2	80	95	95	90	98	91,60
3	Respondent 3	75	75	90	85	75	80,00
4	Respondent 4	50	80	75	90	85	76,00

5	Respondent 5	65	100	90	85	80	84,00
6	Respondent 6	90	95	95	85	80	89,00
7	Respondent 7	75	100	95	100	90	92,00
8	Respondent 8	80	85	95	100	100	92,00
9	Respondent 9	80	100	90	95	100	93,00
10	Respondent 10	80	100	100	100	100	96,00
11	Respondent 11	75	95	98	95	100	92,60
12	Respondent 12	60	100	100	100	100	92,00
13	Respondent 13	90	100	100	95	100	97,00
14	Respondent 14	80	100	100	85	85	90,00
15	Respondent 15	75	75	100	100	100	90,00
16	Respondent 16	50	80	95	100	90	83,00
17	Respondent 17	80	95	100	85	100	92,00
18	Respondent 18	80	100	90	100	75	89,00
19	Respondent 19	90	100	90	80	80	88,00
20	Respondent 20	70	75	85	100	100	86,00
Average Value							88,91

From table 5, is gotten that final value is an average value which is gotten from assessment is carried out by teacher as expert namely 88.91 while the assessment is carried out manually against 20 students which is each of the student doing 5 essay questions that's existing in the system. Value summary which is carried out by expert that represented by teacher and also the gotten value from such kind of system that can be shown on table 6.

Table 6: Expert and system valuation Result

No	Respondent	Value	
		Expert	Jaccard
1	Respondent 1	85,00	78,22
2	Respondent 2	91,60	91,18
3	Respondent 3	80,00	73,03
4	Respondent 4	76,00	72,40
5	Respondent 5	84,00	75,57
6	Respondent 6	89,00	88,50
7	Respondent 7	92,00	89,17
8	Respondent 8	92,00	91,00
9	Respondent 9	93,00	95,00
10	Respondent 10	96,00	95,00
11	Respondent 11	92,60	95,00
12	Respondent 12	92,00	90,00
13	Respondent 13	97,00	95,68
14	Respondent 14	90,00	83,33
15	Respondent 15	90,00	84,67
16	Respondent 16	83,00	79,33
17	Respondent 17	92,00	90,00
18	Respondent 18	89,00	82,78
19	Respondent 19	88,00	86,28
20	Respondent 20	86,00	81,83
Average Value		88,91	85,52

From table 6 is shown the average value from expert is 88.91 and average value which is gotten from the system is 85.52. In here there is

the difference as 3.39 point, with accuration as 96.19%. Can be concluded that average value gotten from the system namely 85.52 is taken from all letters which has similarity among student answer with answer's key, the rest 14.48 is a conjunction, the word mostly existing, and new world that doesn't have answer's key.

Based on the system testing result, so can be concluded that otomatic essay answer assessment system has already been accordance with functional necessity specification which is needed by user.

5. CONCLUSION

The conclusion from this research as follow:

- 1) Using Similarity Jaccard Method that can be impelemented in otomatic essay answer assessment sysem at Al Firdaus Vocational Hight School.
- 2) Accuration is mostly influenced by the availabel key letters that existing in answer and answer's key.
- 3) Assessment of the result of final value is carried out by teacher as expert and assessment is carried out by the system in this research have difference of average value namely 3.39 point, with accuration as 96.19%.
- 4) Similarity Jaccard method can only give rightness value against similar letter in a sentence, and could haven't recognized synonym letter that has similar meaning in the sentence.

Suggestion that can be given from this testing result as follow:

- 1) Using Jaccard Similarity can be supported by another method so as makes possible to correct letter arragement that's existing in tested sentence.
- 2) To complete the similarity jaccard method in order to be noticed about letter synonym, so as if find an answer with similar meaning but using different letter, it can be calculated the similarity value.
- 3) For otomatic essay answer assessment system based on web can be made online so as it's expected can more maximalize in using it.

BIBLIOGRAPHY

- [1]Mardapi, D., 2012. The measurement of education assessment and evaluation. Yogyakarta: Nuha Medika.
- [2]Sudjana, N. (2005) 'Assessment of teaching-learning result process' Bandung: PT Remaja Rosdakarya.
- [3]Arikunto, S. (2012) 'Basic of Education Evaluation second edition', Jakarta: Bumi Aksara.
- [4]Azwar, S., 1996. Achievement Test: Function and development of learning achievement measurement. Yogyakarta: Pustaka Pelajar.
- [5]Jihad, A. and Haris, A., 2008. Study and Learning Essences .
- [6]Rudner, L.M., 2012. Demystifying the GMAT: Computer-Based Testing Terms. Graduate Management Admission Council.
- [7]Feldman, R. and Sanger, J., 2007. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge university press.
- [8]Triawati, C., 2009. Text Mining. Institut Teknologi Telkom. Jawa Barat: Bandung.
- [9]Hayatin, N.,(2015) 'Otomatic Essay Answer Assessment System Based on Sentence Closeness Value', Technology and Manipulation Seminar (SENTRA), pp. 978–979.