Abstract— Essay Scoring is a problem of sentences analysis. We have to be very careful to decide whether two sentences is matched or almost matched or not matched at all. We need to consider several input variables to support matching percentage. This research used number of words, keywords of sentences, and similarity of sentences to decide the concordance of two sentences. Meanwhile, fuzzy logic has tolerance on its input variables using membership function. This characteristic will be useful to decide the approximate percentage of matches. The input variables will be inserted into the fuzzy rule. Then a crisp grade will be the output of Sugeno fuzzy inference. This research will be very useful for lecturers to grade student’s exam on essay test. The result show the proposed essay scoring is reliable to be considered as an alternative for scoring.

I. **INTRODUCTION**

Scoring students’ assessment is a problem for lecturers as it is time consume. It is found that about 30% of teachers’ time is devoted to marking [5]. Nevertheless it is one of important aspect of grading. A good assessment should be capable to evaluate student capability on each lesson indicator.

Evaluation using conventional method is lack of effectiveness and efficiency. It is not effective to use most of the time to check sheet by sheet of paper. Lecturer has to repeat doing the same thing as numbers of students grow. There is also objectiveness problem. Lecturer might score from his/her personal perspective. Not to mention about the paper usage.

For this reason, automated scoring plays an important role to analyze an assessment. The interest in the development and in using of automated assessment systems has grown exponentially in the last few years [2]. Several systems have been developed to solve this problem. There is a method called Latent Semantic Analysis (LSA) observed parameter analysis [4] and LSA with some generalization [2]. Meanwhile, fuzzy logic is suitable to work with perception [1]. several variables are important to support the accuracy of the scoring.

II. **VARIABLE INPUT**

To determine the scoring, data are submitted by students. The writing will be analyzed by using three variable inputs.

A. **Number of words variable**

Number of words is the first input variable. This variable is suitable to grade compactness of precision. Number of words from students’ answer ($\Sigma JM$) and lecturers’ answer ($\Sigma KJ$) will be compared by using (1).

If $\Sigma JM > \Sigma KJ$:

$$vvar1 = 1 - \frac{\Sigma JM - \Sigma KJ}{\Sigma JM}$$

and if $\Sigma JM < \Sigma KJ$:

$$vvar1 = 1 - \frac{\Sigma KJ - \Sigma JM}{\Sigma KJ} \tag{1}$$

B. **Keywords variable**

Keywords are used to anticipate variations of writing for unstructured sentence. The score is not denying main points of the writing. Grading of every keyword depends on priority. The first keyword gets the highest point and reduced as the priority of the next keywords decreased.

$$vvar2 = \frac{\sum \text{weight of matched keywords}}{\sum \text{weight of keywords}} \tag{3}$$

C. **Similarity Variable**

This variable uses similar text algorithm. It counts the number of words or sequence of letters that are the same. The similarity percentage is gained by (4).

$$vvar3 = \frac{\text{number of similar text}}{\text{total length of both text}} + 200 \tag{4}$$
III. **Fuzzy Logic**

A. **Fuzzification**

Fuzzification refers to the process of determining the degree of membership of a crisp input data value among a feature variable’s membership function set. The “fuzzified” values are determined by intersecting the input value to the fuzzy set associated with each linguistic label [3].

Fuzzification is initiated by making sets of fuzzy data from input variables, as shown in Table 1.

Table 1. Fuzzy Data Set

<table>
<thead>
<tr>
<th>Fuzzy Variable</th>
<th>Fuzzy Set</th>
<th>Range</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of words</td>
<td>Equal</td>
<td>[0, 1]</td>
<td>[0.80-1.00]</td>
</tr>
<tr>
<td></td>
<td>Almost equal</td>
<td>[0.50-0.79]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quite equal</td>
<td>[0.10-0.49]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zero</td>
<td>[0]</td>
<td></td>
</tr>
<tr>
<td>keywords</td>
<td>Precise</td>
<td>[0.80-1.00]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Almost precise</td>
<td>[0.50-0.79]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quite precise</td>
<td>[0.10-0.49]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zero</td>
<td>[0]</td>
<td></td>
</tr>
<tr>
<td>similarity</td>
<td>High</td>
<td>[0.80-1.00]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>[0.50-0.79]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>[0.10-0.49]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zero</td>
<td>[0]</td>
<td></td>
</tr>
</tbody>
</table>

B. **Membership function**

The relation between input variables and fuzzification is shown in a membership function. Each member of the set is characterized by its degree of membership within the set [3].

C. **Fuzzy Inference**

Each input variables has 4 members of fuzzy set. There are $4^3$ combinations of rules which is 64 rules. This research use Sugeno zero order (5) to create rules.

If (x1 is A1) • (x2 is A2) • (x3 is A3) • (x4 is A4) • .... •

THEN $z = k$ ........ (5)

X is fuzzy input variables and A is the member of fuzzy set. Both of them create antecedent. Z is output variable and k is the output value. Sample of rules that was created as follow:

[R2] if similar text is HIGH AND number of words is EQUAL AND keywords is QUITE PRECISE THEN grade = 98.5.

[R17] if similar text is MEDIUM AND number of words is EQUAL AND keywords is PRECISE THEN grade = 92.

D. **Defuzzification**

The process of computing a single number that best represents the outcome of the fuzzy set evaluations is called defuzzification [3].

This research used the crisp value of fuzzy inference as the grade value.

IV. **Results and Discussion**

A sample test was made as seen in figures 4.

Apakah yang dimaksud dengan ethernet?

Lectures’ answer:
Ethernet adalah sebuah metode akses media jaringan di mana semua host di jaringan tersebut berbagi bandwidth yang sama dari sebuah link.

Students’ answer:
Ethernet adalah sebuah interface atau sebuah jenis pengkabelan dan pemprosesan sinyal untuk data jaringan.

Maksimum point : 30.
Percentage of correctness : 45%.
Point gained : 13.5
The number of words from students’ answer is 14 and lecturers’ answer is 21 (1).

\[ v_{var1} = 1 - \frac{21 - 14}{21} \]

The percentage for number of words is 66.7%. And from figure 1, we can see that the fuzzy set for number of words is medium.

Keywords analysis for figure 4 is as follow: metode (weight 8), akses (weight 7), media (weight 6), jaringan (weight 5), berbagi (weight 4), bandwidth (weight 3), sama (weight 2), link (weight 1).

\[ v_{var2} = \frac{5}{8 + 7 + ... + 1} \]

The percentage for keywords is 13.9%. The membership function in figure 2 shows that the fuzzy set for it is quite similar.

The next step is to determine similarity of sentences using (4).

\[ v_{var3} = 17.7\% \]

The membership function in figure 3 shows that the fuzzy set is low.

The last step is to find the appropriate fuzzy rule for these three input variables. In this case, the result noted in rule 39, it said:

IF number of words = medium AND keywords = quite similar AND similar text = low THEN grade = 45.

An experiment was set up to determine concordance of lecturer evaluating results and essay scoring results. Writing samples of students were submitted using web interface. The samples consist of 9 students. The teacher grader reviewed each of the 9 writing samples. The same writing samples were analyzed using the scoring system. From table 2, the difference rate is 3.555.

<table>
<thead>
<tr>
<th>NIM</th>
<th>Maximum weight</th>
<th>Lecturer (a)</th>
<th>System (b)</th>
<th>Score difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>42513026</td>
<td>30</td>
<td>10</td>
<td>13.5</td>
<td>3.5</td>
</tr>
<tr>
<td>42513029</td>
<td>30</td>
<td>10</td>
<td>10.5</td>
<td>0.5</td>
</tr>
<tr>
<td>42513031</td>
<td>30</td>
<td>10</td>
<td>13.5</td>
<td>3.5</td>
</tr>
<tr>
<td>42513033</td>
<td>30</td>
<td>15</td>
<td>13.5</td>
<td>1.5</td>
</tr>
<tr>
<td>42513035</td>
<td>30</td>
<td>20</td>
<td>13.5</td>
<td>6.5</td>
</tr>
<tr>
<td>42513037</td>
<td>30</td>
<td>10</td>
<td>0.0</td>
<td>10</td>
</tr>
<tr>
<td>42513039</td>
<td>30</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>42513040</td>
<td>30</td>
<td>20</td>
<td>13.5</td>
<td>6.5</td>
</tr>
<tr>
<td>42513043</td>
<td>30</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
</tr>
</tbody>
</table>

Highest difference 6.5
Lowest difference 0

Errors from essay scoring system were the result of writing errors, foreign terms errors, or abbreviation of words. In the other hand, lecturers‘ scoring denied this kind of error for it could be understood easily by lecturer.

V. CONCLUSION AND FUTURE RESEARCH

The essay scoring can lead to the development of automatic grading systems to help lecturers to minimize inefficiency and lack of objectivity.

The errors that arise in the management of vagueness in the scoring of student sample tests have been discussed. This essay scoring produced average error of 11.85% for the sample test.

The use of three input variables that relies on compactness of precision, main point consideration, and similarity of words are potential variables to be recognized.

Future Research

There are still a lot of variables to be recognized, like synonym or negative statements as well as there are a lot of fuzzy inference method to be developed for better accuracy.

REFERENCES


