SJIF Impact Factor: 6.260| ISI I.F.Value:1.241 Volume: 4 | Issue: 9 | September | 2019 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Peer Reviewed Journal

THE IMPLEMENTATION OF N-GRAM FOR ESSAY EXAM ASSESSMENT

Jesri Hotman Tridayana Purba[1]
[1] *Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia*

Muhammad Zarlis[2]
[2] *Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia*

Sawaluddin[3]
[3] *Department of Mathematics, Universitas Sumatera Utara, Medan, Indonesia*

ABSTRACT
The development of information technology has changed many human lives, including the world of education. At present, computer-based exams begin to be applied on national exams up to the state university entrance exam. Besides the results can be obtained quickly, the computer-based exam method certainly can avoid human mistakes and cheating. But this computer-based written test is still limited to multiple choices. In this paper we introduce the assessment of essay exams that can be performed by computers using the Language Model. The computer will calculate the similarity between the answer key and the answers given by the examinees. The method used is the N-Gram method.

KEYWORDS: *N-Gram, Winnowing, Rolling Hash, Finger Print*

1. INTRODUCTION

Essay test is a kind of learning progress test that requires answers in the form of words. This essay test is very useful for developing the ability to explain or express an opinion in their own language. [1].

Essay tests are divided into two types, they are open essay test and the limited essay test. In an open essay test, the answer is a free or open. This means that students are free to convey the answer with their own descriptions and languages. Meanwhile, in a limited essay test, the answers that are directed and limited. [1].

Open essay assessment will certainly get various answers, and difficult to evaluate. Therefore, in this paper, we use a limited essay test, because it has a standard answer that can be used as a reference or comparison.

2. LANGUAGE MODEL

Language Model is a modeling that will provide a probability value for each word in a sentence or document.[2]

2.1. N-Gram

N-Gram is a Language Model, a method for calculating word order probabilities. N-gram is the simplest method for calculating the probability of word order in a sentence, based on previous words. [2-4]. An N-gram is a sequence of n words. For instance "*Lake Toba is extraordinarily beautiful*", so 2 gram (bigram) contains word order of *"lake toba", "toba is", "is extraordinarily", "extraordinarily beautiful"* or 3 gram (trigram) may contain word order of *"lake toba is", "toba is extraordinarily, " is extraordinarily beautiful.* Likewise 4 grams will contain a sequence of 4 words.

The N-gram method can also be applied in alphabetical order, for example "*Lake Toba is*

extraordinarily beautiful", then normalized into "*laketobaisextraordinarilybeautiful*", so the 4-gram series is: *lake, aket, keto, etob, toba, obai, bais, aise, isex, sext, extr, xtra, trao, raor, aord, ordi, rdin, dina, inar, nari, aril, rily, ilyb, lybe, ybea, beau, eaut, auti, utif, tifu, iful.*

N-grams can be used to calculate the probability of the word (w) appearing in a sentence (h). This can be denoted as P (w | h). For example, in the sentence *'lake toba is extraordinarily'*, the probability of the word 'beautiful' appearing as the next word can be denoted as:

P (beautiful| lake toba is extraordinarily) (1)

To calculate the probability above, it is done by counting the number of occurrences of the phrase 'lake toba is extraordinarily' in a large corpus, followed by the word 'beautiful', as follows:

$$P(\text{beautiful} \mid \text{lake toba is extraordinarily}) = \frac{C(\text{lake toba is extraordinarily beautiful})}{C(\text{lake toba is extraordinarily})} \quad (2)$$

2.2. Winnowing Algorithm

The Winnowing algorithm processes every text in this n-gram sequence, converting it into a set of hash values. This collection of hash values is called a fingerprint, which will be used to compare text similarities [5-8]. The steps of the Winnowing Algorithm are as follows:

Step 1. Text normalization. In this step all capital letters will be converted to lowercase letters, all spaces, punctuation marks and special characters will be removed. The sentence "Lake Toba is extraordinarily beautiful" is normalized to: "laketobaisextraordinarilybeautiful"

Step 2. N-gram Tokenization. In this step, the normalized text is formed into a series of n-grams or referred to as n-gram tokenization. In this case, the n = 4, so the 4-gram tokenation of the text "laketobaisextraordinarilybeautiful" is:

lake aket keto etob toba obai bais aise
isex sext extr xtra trao raor aord ordi rdin
dina inar nari aril rily ilyb lybe ybea beau
eaut auti utif tifu iful

Step 3. Calculating Hash Value. The hash value is calculated using the Rolling Hash Method. This method is stated by Rabin Karp, where the hash function H (c1 ... ck) is defined as equation (3) follows [3]:

$$H_{(c1 \ldots ck)} = c_1 * b^{(k-1)} + c_2 * b^{(k-2)} + \ldots + c_{k-1} * b + ck \quad (3)$$

where,

c: ASCII character value
b: Base prime numbers
k: n-gram value

for the example above we get the hash value:

4211 4001 4257 4202 4522 4275
3949 4010 4293 4490 4269 4723
4560 4398 4060 4428 4403 4072
4230 4290 4068 4468 4268 4400
4549 3963 4067 4125 4620 4500
4212

Step 4. Forming Window. The windows process is the same as the n-gram process, in this case it is formed from the hash values above. The hash values obtained are formed into hash columns. If the window is 4, then a series of hashes in the form of a window is formed as follows:

4211 4001 4257 4202
4001 4257 4202 4522
4257 4202 4522 4275
. . . .
. . . .
. . . .
3963 4067 4125 4620
4067 4125 4620 4500
4125 4620 4500 4212

Step 5. Setting fingerprint. From the window formed, then we can seek for the smallest hash value of each row. This value is called a fingerprint. If there is a smaller hash value of more than 1, then the rightmost hash value is taken. From the example above we get a fingerprint: 4001, 4202, 3949, 4010, 4269, 4060, 4072, 4068, 4268, 3963, 4067 and 4125.

Step 6. Calculate the similarity. To calculate the similarity, it uses Jaccard's Similarity Coeficient, by the equation formula (4) :

$$D(A,B) = \frac{|A \cap B|}{|A \cup B|} x\ 100 = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} x\ 100\% \quad (4)$$

If the sentence A = "Lake Toba is extraordinarily beautiful" and B = "Lake Toba beautiful", then the Winnowing steps obtained are:

Fingerprint A : **4001 4202** 3949 4010 4269 4060 4072 4068 4268 **3963 4067 4125** = 12

Fingerprint B : 4001 4202 3914 3901 3963 4067 4125 = 7

and |A∩B| = |4001 4202 3963 4067 4125| **=** 5

So :

$$D(A,B) = \frac{5}{12+7-5} x\ 100\% = 35.71\%$$

3. PROPOSED METHOD

To improve the accuracy of assessment, the answer key is considered as training data, so the answer key is a collection of *some possible answers* that may be given by students, with different words (synonyms), but the meaning is the same. With 'Some Possible Answer' method, although the answers from students vary, differ in words or sentences, but still have the same meaning, can still be assessed.

For example: *Mention the meaning of renewable energy.*

Then, there are several answers, where the sentences are different but basically the meaning is the same, and can be considered correct, namely:
Answer 1. *Renewable energy is the energy which is derived from "sustainable natural processes"*
Answer 2. *Renewable energy is energy that comes from nature and can be continuously produced.*
Answer 3. *Renewable energy is energy produced from natural sources that will not run out.*
So these three answers will be used as training data (answer key), and as a counterweight, each similarity value will be multiplied by the number of *Some Possible Answers* (n) in the training data (answer key), so it is obtained from equation (4):

$$D(A,B) = \frac{|A \cap B|}{|A|+|B|-|A \cap B|} x\ n\ x\ 100\% \qquad (5)$$

4. RESULTS AND DISCUSSION

4.1 Results

The study was conducted on 28 students, with the number of questions 5, so it has 140 pieces of data. Then, there are 2 types of answer keys, namely keys that only provide 1 answer, and keys that provide *some possible answers*. Then the score is converted into an index value with conditions as in table 1.

Table 1. Converting Score to Index

Index	Score
A	Score >=80
B	79>= Score>=68
C	67>= Score >=56
D	55>= Score>=45
E	Score <= 44

From the tests conducted, the results are obtained as in table 2 below:

Table 2. Test result

No	Student ID	Question Number	Teacher Scores		Score by Computer					
					1 Answer Key			Some possible answers		
			Score	Index	Score	Index	Correct/ Wrong	Score	Index	Correct/ Wrong
1	1	1	100.00	A	100.00	A	1	100.00	A	1
2	1	2	80.00	A	15.15	E	0	100.00	A	1
3	1	3	100.00	A	100.00	A	1	100.00	A	1
4	1	4	100.00	A	3.45	E	0	100.00	A	1
5	1	5	-	E	-	E	1	18.18	E	1
6	2	1	100.00	A	100.00	A	1	100.00	A	1
7	2	2	50.00	D	7.69	E	0	100.00	A	0
8	2	3	100.00	A	100.00	A	1	100.00	A	1
9	2	4	100.00	A	5.56	E	0	100.00	A	1
10	2	5	-	E	-	E	1	15.87	E	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
138	28	3	100.00	A	100.00	A	1	100.00	A	1
139	28	4	100.00	A	-	E	0	100.00	A	1
140	28	5	80.00	A	18.92	E	0	94.74	A	1
					Correct Answer		86			121
					Percentage		61.43%			86.43%

1:Correct, 0:Wrong

4.2 Discussion

From the Table 2 of the test results, 3 values are presented, they are the value given by the teacher, the value given by the computer from 1 answer key, and the value given by the computer from *some possible answers*. From the results obtained, it appears that with

1 key answer, the total correct is 86 or 61.43%, while with *some possible answers*, the total correct is 121 or 86.43%. Visually it can be seen in Figure 1.

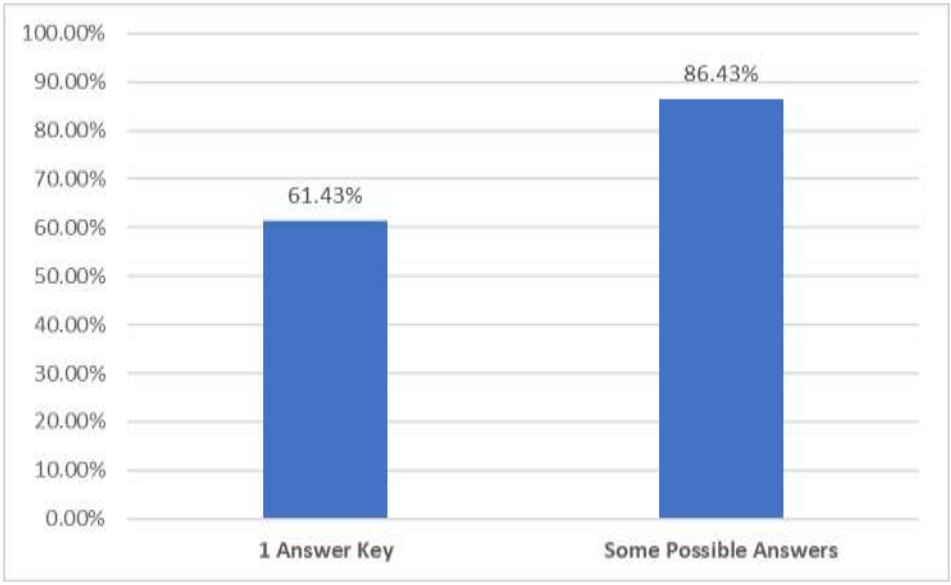

Figure 1. Graph of accuracy between *1 key answer* and *Some Possible Answers*

To find out more, presented a graph of the test results with *1 answer key* and with *some possible answers* for each number of questions. The graph can be seen in Figure 2.

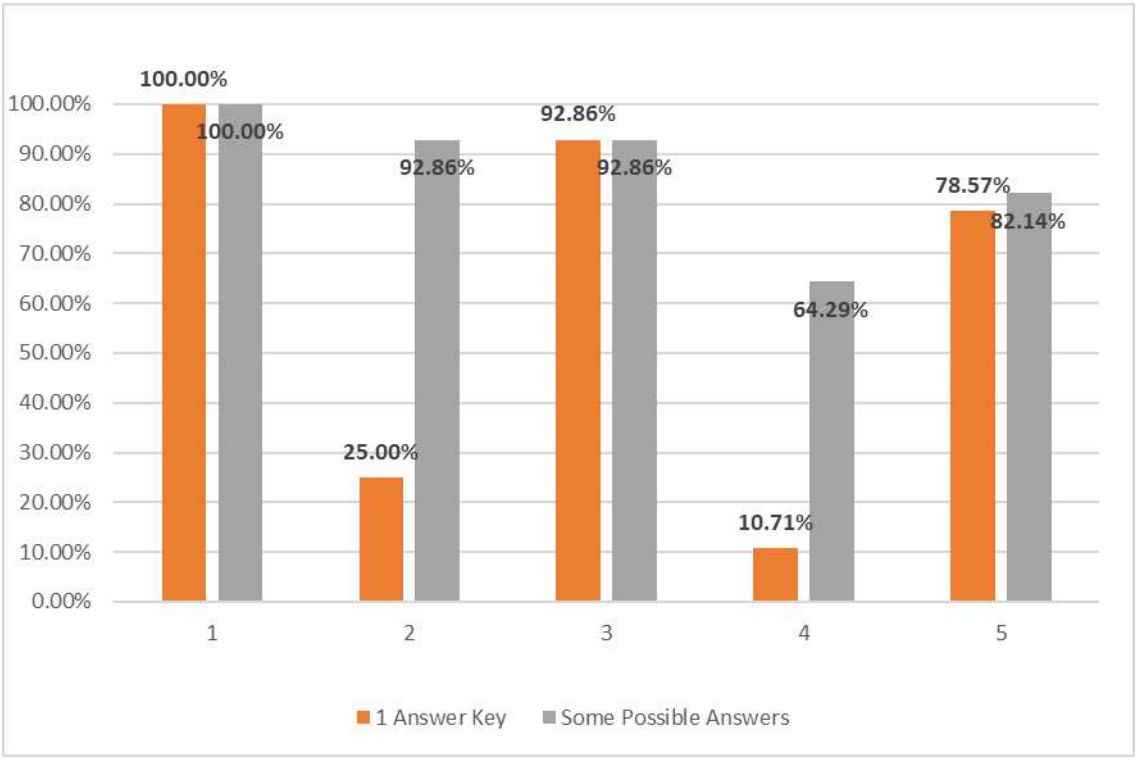

Figure 2. Test Result Graph per Question Number

For question number 1 and question number 3, you can see that the accuracy value is the same between *1 answer key* and *some possible answers*. This is because the answers to questions number 1 and 3 are certain answers. While questions number 2, 4, 5 are questions that ask students to explain something, so that each student's answers can vary. With proposed method '*some possible answers*' , the accuracy can be improved.

5. CONCLUSIONS AND SUGGESTIONS

5.1 Conclusions

From the series of tests and discussions above it can be concluded that:

1. N-Gram can be used for essay examination assessment.
2. The accuracy of the assessment is influenced by the type of questions according to student's desired answer.
3. Because it is calculating the similarity, this method is only effective for limited essay tests.
4. To improve the accuracy of the assessment, *some possible answers* are needed as key answers.

5.2. Suggestions

The advice that can be given in further research is the use of more types of questions. Research is also suggested to determine possible answers, whether by creating a database of synonym words so that better results will be obtained.

REFERENCES

1. *Ngalimun 2018 Evaluasi dan Penilaian Pembelajaran (Yogyakarta: Penerbit Parama Ilmu) pp 95-97*
2. *Jurafsky D and Martin J H 2017 Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (New Jersey: Prentice Hall) p 36*
3. *Darujati, C., Gumelar, A.B. 2012. Pemanfaatan Teknik Supervised untuk Klasifikasi Teks Bahasa Indonesia. Jurnal Link 16(1): 5-1 – 5-8.*
4. *Rasywir, E., Purwarianti, A. 2015. Eksperimen pada Sistem Kalsifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin. Jurnal Cybermatika 3 (2) : Artikel 1.*
5. *Kharisman O, Susanto B and Suwarno S 2013 Implementasi Algoritma Winnowing untuk Mendeteksi Kemiripan Pada Dokumen Teks. Informatika Vol 9, No. 1, April 2017.*
6. *Purwitasari D, Kusmawan P Y and Yuhana U L 2011 Jurnal Ilmiah Kursor Deteksi Keberadaan Kalimat Sama Sebagai Indikasi Penjiplakan Dengan Algoritma Hashing Berbasis N-Gram V 6 pp 37-44*
7. *Suyanto. 2017. Data Mining untuk Kalsifikasi dan Klasterisasi Data. Informatika Bandung: Bandung.*
8. *Prasetyo, E. 2014. Data Mining – Mengolah Data menjadi Informasi Menggunakan Matlab. Penerbit Andi: Yogyakarta.*