TEXT SUMMARIZING SYSTEM OF ENGLISH SUBJECTS AND TEXT MINING SUBJECTS FOR COMPUTER SCIENCE STUDENTS

Aa Zezen Zaenal Abidin[1], Yuli Murdianingsih[2], Usep Tatang Suryadi[1], Didik Setiyadi[3]

[1]Informatics Engineering Department, STMIK Subang, Indonesia
[2]Informatics Management Department, STMIK Subang, Indonesia
[3]Informatics Engineering Department, Bina Insani University, Indonesia

Abstract
The development of text data is the birth of text mining, one of the techniques of managing text in order to summarize text is called text summarizer. in this study the text summarization uses the TF-IDF method. Text summarizing system is used in increasing students' interest in learning English, especially the ability to read news in English, where learning English in Indonesia experiences obstacles at various levels of formal education, from elementary schools, junior high schools, senior high schools to tertiary institutions. Text summarizer have created, used to demonstration method in English and text mining subject in departemant of informatics. Text summarizing system is also used as a demonstration method when entering text summary lecture material as a sub subject of text mining course material. stemming process, word tabulation, idf calculation and determination of word weight and sentence weight, the system is summarized with a summary of activities performed by a number of twenty English teachers and testing by ten online text summaries. Calculated the value of recall, the value of precision and f-measure. Text summarizing system is used in English learning activities as a medium of learning in order to present learning motivation, and demonstrated in the text mining course during the initial lecture on sub text summarizing material. Testing system by twenty English teachers that Text Summarizing System using TF-IDF in this study obtained a recall value of 0,687 precision value of 1,00 and f-measure test value of 0,814. Testing By online text summarizing systems using TF-IDF, in this study obtained a recall value of 0,675, precision value of 1,00 and f-measure test value of 0,8059 which means that the summarizing system has reliability and can be used as a medium for learning English courses and text mining. The future research is determining the steps to use text summarizing media and evaluating the benefits of learning optimization and also assessing perceptions from lecturers as well as students on more respondens.

Keywords: Text mining, Tf-Idf, English teaching, Text summary, Online Text Summary

INTRODUCTION
There are several issues in Indonesia as regards learning English. English learning especially in formal schools such as primary, junior high school, senior high schools and university. There difficulties in the conduct of studying English in elementary Schools In Indonesia, The First, the teacher still uses approaches that make students uninterested in embracing the instructional material, both schools do not promote English learning and third, the teacher refuses to incorporate English in the curriculum [1].

There are four problems in junior high scholls[2] explained as follows:

1) The first question lies in students 'passivity in the English learning process. Some students find English to be a scary language, so when they want to ask something, other questions will appear first in the form of "What do I say" or "What do you call it in English."
2) The second, the second question, the mastery of English vocabulary by students is still limited, while the relevant curriculum allows English teachers to use English during the process of teaching and learning. While, Teacher Talk must be driven essentially by the English language used by the teacher, in the sense that English spoken by the teacher must be simple and clear. But what if a teacher's pronouncement is isolated from the Teacher Talk concept? This would of course make students more confused, because the words of the instructor will be too loud.
3) The third problem concerns applying the Empirical Method. Indeed, we can already infer from the name alone that when used in teaching mathematics and natural sciences and its branches of science, this approach appears to be more fitting. As far as its application in teaching linguistics is concerned, it appears to be complicated and to be very pushy.
4) The fourth question relates to teachers 'failure to grasp the implementation plan for the 2013 Curriculum

According to [3], there are sixtenh problem in teaching English in senior high scholls:

1) Teachers did not undergo sufficient training on Kurtilas;
2) Power Books not available to teachers and students;
3) Inadequate Facilities & Targets available;
4) Teachers Not yet Able to Use Information Technology (IT);
5) Difficulty of teachers in planning 2013 RPP based curriculum;
6) Teachers experiencing difficulties in performing tests based on a program in 2013;
7) The inaccessibility of 2013 Professional Standards for Curriculum-based Assessment;
8) The teacher faces difficulties in improving the way students are learning by passivity;
9) The teacher has trouble applying the five steps of the scientific method (5 M) The teacher has not undergone sufficient training on Kurtilas;
10) The teacher still has no specific guidelines on models and monitoring methods for student learning results based on Kuri-kulum 2013;
11) There are teachers who lack "place" because most schools do not allow Language Specialization;

12) The lack of available hours for meeting the prescribed requirements;
13) Lack of instructor skills in choosing and using tools for learning;
14) Instructions on how to assign subject matter by semester are not yet available;
15) The teacher is having difficulty teaching students to use high order thinking skills (HOTS);
16) The matrices system for the students is not yet scheduled.

According to students Speaking is at the highest level, some factors that are the reason why they prefer Speaking are lack of English as speaking, listening, writing and reading[4], can be explained as follows:

1) The first, speaking is at the highest level, the next level is listening, writing and the lowest is reading, Difficult speaking because it is very different from Indonesian, fear of making errors, fear of being laughed at by peers, and lack of grammar skills, factors that make it difficult to learn English.
2) The second issue is listening skills, students feel unable to keep up with the usual pace of native English speakers while communication to English language videos or shows, Then the lack of vocabulary mastery and comprehension of the English accent leaves them unable to understand the substance of the conversation, even if the pace was set to Indonesian or the speakers were not native.
3) The third is Writing difficulty is the most challenging, few students find writing difficult to learn as they follow the writing process where there are many phases to pass before publishing their English work, the writing process consists of planning, writing, reviewing, revising, and publishing.
4) The lowest difficulty in learning English lies in Reading, Most students think that reading skills are the easiest thing to do, this is because students have an interest in reading activities, so even though the language of instruction given is English they still enjoy the activity, the second reason is that when reading students have text that can be directly used as reference material to answer questions related to understanding the text.

Problems of studying English in a variety of formal education definitely influence the results of formal education for the general public. One of the things that the general public is doing is reading the news. While reading ability has been said to be the lowest level problem, in reality it is uncommon among students who read news in English because of English ability.

Text Mining, this course is a compulsory subject with a load of 2 SKS, a group of subjects working on language processing methods in the form of unstructured text data processing. As for the purpose of its implementation is the Student is able to recognize various kinds of text mining applications and Students able to implement one of the text mining systems using text processing methods. As for some of the study material is an introduction to text processing and text mining, demonstration of text mining applications, preprocessing, stopword list and stemming, feature selection (using methods commonly used in text mining) both statistical and linguistics, information retrieval systems, Document Classification, Document Clustering, Text Summarization, Information Extraction, Use of PHP and DBMS Programming Language Functions in Making Text Mining Applications and Final Projects.

To build a sense of belonging in the field of computer science, the exposure of the lecture material held must contribute to each other, both in the field of special expertise and general expertise such as English. text summarizing system is used for learning media in the ability to resume English language content including news in English. Students get a picture of the use of computer systems in resumes as well as having inspiration for system development because of their critical attitude in comparing the capabilities of machines and humans. as one of the uses of text mining in the world of education[5], especially summarizing news texts in English, although there are already many online text summarizing systems, there are always obstacles such as some that are paid, sometimes the hassle of having to fill in biodata is also sometimes a problem of communication signals, so that it feels long even though it only resumes short sentences in learning activities.

Demonstration is one method of learning. especially courses that must present the ability of students to create one of the fields of text mining in the project, the demonstration of the text summarizing system at the beginning is expected to be very helpful in giving a picture and spirit of students to be able to complete the learning process completely.

Therefore, it is expected that the use of a text summarizing media system in English and text mining can be an effort to optimize the integration of the course administration, contributing to the reduction of the problem of English language learning in higher education. the use of demonstration methods at the beginning of the discussion of summarizing text with the entire process in the implementation of the method can be helpful for students. course content is not only theory or concept, but the concept is accompanied by a full presentation of the implementation, so it is expected to be an example for students in the learning process, which shows the lecturer also loves the contents of his lecture in full.

MATERIALS AND METHOD
The first time a text summarizing system was made for the English language news using the tf-idf method. English news text summarizing system is tested to obtain the value of recall, precision and f-measure, text summarizer used to teach activity in English subject and Text Mining subject shown in figure 1. Preprosessing step will be syntax and semantical analysis, in the form of stop word removing process, text transformation stage will be done stopword list process or stemming process, feature selection stage is done by counting the number of words that appear using simple statistics or related methods, using TF method -IDF is done by calculating the weight of each word in a sentence, calculating the sum of the weight of each word so that it becomes the weight of the sentence, determining the sentences with the highest weight will be selected and also limiting the maximum number of sentences summarized, so that a summary is obtained. Because the system will be used for learning media in English language courses and text mining, testing is needed by two testers, namely English language teachers in schools and by other existing text summarizing systems and online so that empirically adequate reliability is obtained.
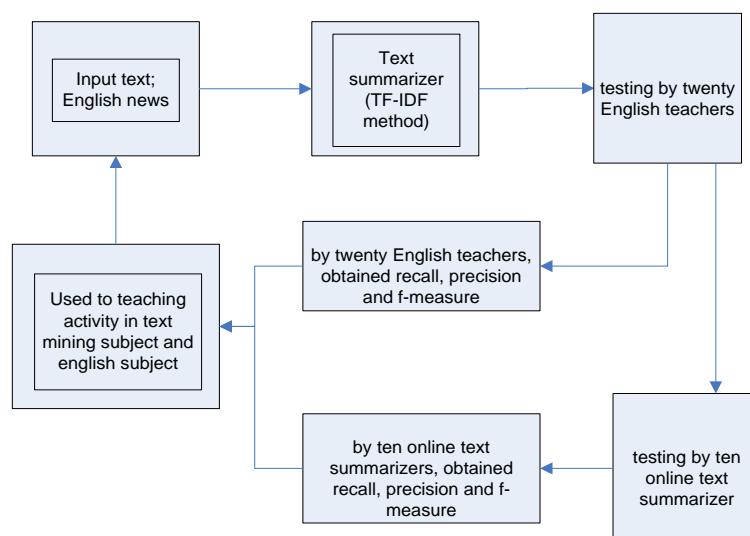
Figure 1 Research Activity

Then reading the English daily news is one of them for developing language skills. Many online media exist as an alternative in English as follows:

1. ABC News
2. AL Arabiya
3. Al Jazeera
4. CNN News
5. Express
6. Euronews
7. Fox News
8. Jakarta Post
9. New York Times
10. Muslim News

there are somes text summary online can use in teaching activity:

1. Tools4noobs
2. Resoomer
3. Textcomparator
4. SMMRY
5. freesummarixer
6. Summarygenerator
7. simplish
8. Summarizethis
9. Summarizethis
10. Autosummarixer

Text Mining is the method of collecting data in the form of unstructured text, with some of the techniques used in text mining such as natural language processing, text classification and grouping, information retrieval and text summarization[5]. Makes online knowledge extraction, news summary in Spanish, a program that tracks news sources and summarizes the latest published articles, the ability to summarize user-furnished texts in English and Spanish, shows the versatility of event extraction as it recognizes specific concepts that define different types of events by mining English text material[6]. Automatic text summation is an application that aims to shorten the source text to a shorter edition [7].

In this analysis, use of the TF-IDF method is the approach used by the authors to summarize automated texts in English. This method is achieved by assigning the document (D) the weight of a word's (t) reference. Every word is called a text, for a single language. This approach uses two principles for calculating weight, namely the term frequency (TF) and the inverse frequency of the text (IDF). Term frequency (TF) is the number of times a word (t) in sentence (D) appears. Frequency of Inverse Documents (IDF) is determined using frequency of document (df). The length of the text is the number of sentences (D), containing the word (t).

In this method word weighting in a text is achieved by multiplying the TF and IDF values. TF means as frequency of term[8], A term's weight is greater if the word regularly appears in a text, and less if the term appears in less documents, Some of the measurements used in the TF-IDF process include weighting of TF-IDF .

$$IDF = \log(\frac{N}{df_i}) \dots\dots\dots\dots\dots\dots(1)$$

Where,

N   = number of sentences with a phrase (t)

Idf   = number of occurrences of terms (duration) in respect of D

Calculate weight [8] (W)

$$TF - IDF = W_{d.t} = TF_{d.t} * IDF_t \dots\dots\dots\dots\dots\dots(2)$$

Where:

D   = sentence to-D

t   = term to-t

TF   = term frequency

W   = wight of sentence to-D against term to-t

IDF   = invers document frequency

For evaluation text summary system, using f-measure, according to[9][10] [11]

$$f - measure = \frac{2 \: x \: precision \: x \: recall}{precision + recall} \dots\dots\dots\dots 3$$

The steps to work on automatic text summarization using the TF-IDF method performed in this study are as follows:

1) First, Count the number of sentences (D) in English documents.
2) Second, the process of eliminating common words (Stopword) to reduce the number of occurrences of words that have no meaning.
3) Third, the process of returning a word (stemming) contained in a sentence (D) into the form of words (t) basic.
4) Fourth, Calculate the term frequency (tf), which is the number of times the word appears (t) in the sentence (D).
5) Fifth, Calculate document frequency (df), i.e. the number of sentence frequencies (D) containing the word (t).
6) Sixth, Calculate the inverse document frequency (IDF) by logging the total number of sentences (D) in English documents divided by document frequency (df).
7) Seventh, Calculate (W) the weight of each word (t) in a sentence (D) by multiplying the term frequency (tf) with the inverse document frequency (IDF).
8) Eighth, calculate the total weight of sentences (D) by adding up (W) the weights of each word (t) in sentences (D).
9) Ninth, Based on the results of the calculation of the total weight of sentences (D), 40% will be taken of the total sentence weight (D) which has the highest value.
10) Tenth, From the steps - steps then produce a collection of sentences (D) which contains important parts of English documents.

The source of English news is from The Jakarta Post:

The House of Representatives is working to ensure that everything is well-organized for the visit of King Salman bin Abdulaziz al-Saud of Saudia Arabia.

The House's secretariat general said the legislative body had coordinated with the Saudi kingdom's protocol affairs office to find out all the things King Salman might need during his visit.

The secretariat general further said that the House had started to beautify the House complex for the king. King Salman will visit Indonesia from March 1 to March 9.

Damayanti, the secretariat general's deputy chairman for hearings, said physical preparations to welcome King Salman would take place over two weeks.

To assist the king, the House will install a portable ramp in front of the Nusantara Building. Built in 1968, certain parts of the Nusantara Building are not friendly to the elderly and people with disabilities. King Salman is 81 years old.

She further explained that King Salman also required clean toilets. King Salman will use a special toilet similar to the one used by the President. A room containing medical equipment has also been prepared, she added.

Step 1. Count the number of sentences (D) in English documents.

1) The House of Representatives is working to ensure that everything is well-organized for the visit of King Salman bin Abdulaziz al-Saud of Saudia Arabia.
2) The House's secretariat general said the legislative body had coordinated with the Saudi kingdom's protocol affairs office to find out all the things King Salman might need during his visit.
3) The secretariat general further said that the House had started to beautify the House complex for the king.
4) King Salman will visit Indonesia from March 1 to March 9.
5) Damayanti, the secretariat general's deputy chairman, said physical preparations to welcome King Salman would take place over two weeks.
6) To assist the king, the House will install a portable ramp in front of the Nusantara Building.
7) Built in 1968, certain parts of the Nusantara Building are unfriendly to the elderly and people with disabilities.

8) King Salman is 81 years old.
9) the secretariat general's deputy chairman, further explained that King Salman also required clean toilets.
10) King Salman will use a special toilet similar to the one used by the President.A room containing medical equipment has also been prepared, she added.

Step 2. The process of eliminating common words (Stopword) to reduce the number of occurrences of words that have no meaning.

1. **The** House **of** Representatives **is** working **to** ensure **that** everything **is** well-organized **for the** visit **of** King Salman bin Abdulaziz al-Saud **of** Saudia Arabia.
2. **The** House's secretariat general said **the** legislative body **had** coordinated **with the** Saudi kingdom's protocol affairs office **to** find out **all the things** King Salman might need during **his** visit.
3. **The** secretariat general further said **that the** House **had** started **to** beautify **the** House complex **for the** king.
4. King Salman **will** visit Indonesia **from** March 1 **to** March 9.
5. Damayanti, **the** secretariat general's deputy chairman, said physical preparations **to** welcome King Salman **would** take place over two weeks.
6. **To** assist **the** king**, the** House **will** install **a** portable ramp **in front of the** Nusantara Building.
7. Built **in 1968**, certain parts **of the** Nusantara Building **are** unfriendly **to the** elderly **and** people **with** disabilities.
8. King Salman **is 81** years old.
9. **the** secretariat general's deputy chairman, further explained **that** King Salman **also** required clean toilets.
10. King Salman **will** use **a** special toilet similar **to the** one used **by the** President.
11. **A** room containing medical equipment **has also been** prepared, **she** added.

Then the results were:

1. House Representatives working ensure everything well-organized   visit King Salman bin Abdulaziz al-Saud Saudia Arabia.
2. House's secretariat general said legislative body coordinated with Saudi kingdom's protocol affairs office find out King Salman might need during visit.
3. secretariat general further said House started beautify House complex king.
4. King Salman visit Indonesia March 1 March 9.
5. Damayanti, secretariat general's deputy chairman, said physical preparations welcome King Salman take place over two weeks.
6. assist king, House install portable ramp Nusantara Building.
7. Built , certain parts Nusantara Building unfriendly elderly people with disabilities.
8. King Salman years old.
9. secretariat general's deputy chairman, further explained King Salman required clean toilets.
10. King Salman use special toilet similar one used President.
11. room containing medical equipment prepared, she added.

Step 3. stemming process:

1. House **Representatives working** ensure everything well-**organized**   visit King Salman bin Abdulaziz al-Saud Saudia Arabia.
2. **House's** secretariat general **said** legislative body coordinated Saudi **kingdom's** protocol **affairs** office find out King Salman **might** need during visit.
3. secretariat general further **said** House **started beautify** House complex king.
4. King Salman visit Indonesia March 1 March 9.

5. Damayanti, secretariat **general's** deputy chairman, **said** physical **preparations** welcome King Salman take place over two **weeks**.
6. assist king, House install portable ramp Nusantara **Building**.
7. Built 1968, certain **parts** Nusantara **Building** unfriendly elderly people with **disabilities**.
8. King Salman 81 **years** old.
9. secretariat **general's** deputy chairman, further **explained** King Salman **required** clean **toilets.**
10. King Salman use special toilet similar one **used** President.
11. room **containing** medical equipment **prepared, added.**

The results was :

1. House Represent work ensure everything well-organize visit King Salman Saudia Arabia.
2. House secretariat general say legislative body coordinate Saudi kingdom protocol affair office find out King Salman may need during visit.
3. secretariat general further say House start beauty House complex king.
4. King Salman visit Indonesia March 1 March 9.
5. Damayanti, secretariat general deputy chairman, hear say physical prepare welcome King Salman take place over two week.
6. assist king, House install portable ramp Nusantara Build.

7. Built, certain part Nusantara Build unfriendly elderly people with disability.
8. King Salman year old.
9. secretariat general deputy chairman, further explain King Salman require clean toilet.
10. King Salman use special toilet similar one use President.
11. room contain medical equipment prepare, add.

Step 4. Calculate the term frequency (tf), which is the number of times the word (t) appears in the sentence (D).

Step 5. Calculate document frequency (df), i.e. the number of sentence frequencies (D) containing the word (t).

While the IDF, inverse document frequency calculated for examples as shown in process as follows:

$$IDF=\log(\frac{11}{4}) =0,439$$

$$IDF=\log(\frac{11}{1}) =1,0439$$

and so on as shown in the table 1, recapitulations of all IDF calculation from term to-1 to 70.

Table 1 Recapitulations of All IDF Calculation from Term to- 1 to 70

No	term	Tf												idf
		1	2	3	4	5	6	7	8	9	10	11	df	log(n/df)
1	house	1	1	1			1						4	0,439333
2	representative	1											1	1,041393
3	work	1											1	1,041393
4	ensure	1											1	1,041393
5	everything	1											1	1,041393
6	well	1											1	1,041393
7	organize	1											1	1,041393
8	visit	1	1		1								3	0,564271
9	king	1	1	1	1	1	1		1	1	1		9	0,08715
10	saudi	1	1										2	0,740363
11	arabia	1											1	1,041393
12	secretariat		1	1		1				1			4	0,439333
13	general		1	1		1				1			4	0,439333
14	say		1	1		1							3	0,564271
15	legislative		1										1	1,041393
16	body		1										1	1,041393
17	coordinate		1										1	1,041393
18	kingdom		1										1	1,041393
19	protocol		1										1	1,041393
20	affair		1										1	1,041393
21	office		1										1	1,041393
22	find		1										1	1,041393
23	out		1										1	1,041393
24	thing		1										1	1,041393

No	Term												Total	Value
25	may	1											1	1,041393
26	need	1											1	1,041393
27	during	1											1	1,041393
28	further		1							1			2	0,740363
29	beautify		1										1	1,041393
30	complex		1										1	1,041393
31	indonesia			1									1	1,041393
32	march			1									1	1,041393
33	deputy				1					1			2	0,740363
34	chairman				1					1			2	0,740363
35	hear				1								1	1,041393
36	physical				1								1	1,041393
37	prepare				1							1	2	0,740363
38	welcome				1								1	1,041393
39	take				1								1	1,041393
40	place				1								1	1,041393
41	over				1								1	1,041393
42	two				1								1	1,041393
43	week				1								1	1,041393
44	assist					1							1	1,041393
45	install					1							1	1,041393
46	portable					1							1	1,041393
47	ramp					1							1	1,041393
48	front					1							1	1,041393
49	nusantara					1	1						2	0,740363
50	build					1	2						2	0,740363
51	certain							1					1	1,041393
52	part							1					1	1,041393
53	unfriendly							1					1	1,041393
54	elderly							1					1	1,041393
55	people							1					1	1,041393
56	disabilities							1					1	1,041393
57	years								1				1	1,041393
58	old								1				1	1,041393
59	explain									1			1	1,041393
60	require									1			1	1,041393
61	clean									1			1	1,041393
62	toilet									1	1		2	0,740363
63	use										2		1	1,041393
64	special										1		1	1,041393
65	similar										1		1	1,041393
66	president										1		1	1,041393
67	room											1	1	1,041393

68	contain		1	1	1,041393
69	medical		1	1	1,041393
70	equipment		1	1	1,041393

the document frequency (df).

Step 7. Calculate the weight of each word (t) in a sentence (D) by multiplying the term frequency (tf) with the inverse document frequency (IDF).

Step 8. Calculate the total weight of the sentence (D) by adding up (W) the weight of each word (t) in the sentence (D).

The word weight and the term weight are determined from the eleven sentences. Multiplication of existing TF from and IDF from table 1, values for each word, by word weight, out of 70 words or terms, as shown in table 2. Four sentences with the highest sentence weight weights are taken as summaries. The first, second, sixth and seventh sentences are four phrases with the greatest weight. The meaning of each is the first sentence of ..., the second sentence of .... And the seventh term ....

$$W_{1.1} = TF_{1.1} * IDF_1 = 1 * 0,439333 = 0,439333$$

$$W_{1.2} = TF_{1.2} * IDF_2 = 1 * 0,439333 = 0,439333$$

$$W_{1.3} = TF_{1.3} * IDF_3 = 1 * 0,439333 = 0,439333$$

$$W_{1.6} = TF_{1.6} * IDF_t = 1 * 0,439333 = 0,439333$$

and so on as shown in the table 2, recapitulations of all wight calculation from term to-1 to 70.

Tabel 2 Recapitulation of All wdt Calculation for Term to-1 to 70

idf	wdt=tf*idf										
log(n/df)	1	2	3	4	5	6	7	8	9	10	11
0,439333	0,44	0,44	0,44	0	0	0,44	0	0	0	0	0
1,041393	1,04	0	0	0	0	0	0	0	0	0	0
1,041393	1,04	0	0	0	0	0	0	0	0	0	0
1,041393	1,04	0	0	0	0	0	0	0	0	0	0
1,041393	1,04	0	0	0	0	0	0	0	0	0	0
1,041393	1,04	0	0	0	0	0	0	0	0	0	0
1,041393	1,04	0	0	0	0	0	0	0	0	0	0
0,564271	0,56	0,56	0	0,56	0	0	0	0	0	0	0
0,08715	0,09	0,09	0,09	0,09	0,09	0,09	0	0,09	0,087	0,09	0
0,740363	0,74	0,74	0	0	0	0	0	0	0	0	0
1,041393	1,04	0	0	0	0	0	0	0	0	0	0
0,439333	0	0,44	0,44	0	0,44	0	0	0	0,439	0	0
0,439333	0	0,44	0,44	0	0,44	0	0	0	0,439	0	0
0,564271	0	0,56	0,56	0	0,56	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0

1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
1,041393	0	1,04	0	0	0	0	0	0	0	0	0
0,740363	0	0	0,74	0	0	0	0	0	0,74	0	0
1,041393	0	0	1,04	0	0	0	0	0	0	0	0
1,041393	0	0	1,04	0	0	0	0	0	0	0	0
1,041393	0	0	0	1,04	0	0	0	0	0	0	0
1,041393	0	0	0	1,04	0	0	0	0	0	0	0
1,041393	0	0	0	1,04	0	0	0	0	0	0	0
1,041393	0	0	0	1,04	0	0	0	0	0	0	0
0,740363	0	0	0	0	0,74	0	0	0	0,74	0	0
0,740363	0	0	0	0	0,74	0	0	0	0,74	0	0
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
0,740363	0	0	0	0	0,74	0	0	0	0	0	0,74
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
1,041393	0	0	0	0	1,04	0	0	0	0	0	0
1,041393	0	0	0	0	0	1,04	0	0	0	0	0
1,041393	0	0	0	0	0	1,04	0	0	0	0	0
1,041393	0	0	0	0	0	1,04	0	0	0	0	0
1,041393	0	0	0	0	0	1,04	0	0	0	0	0
1,041393	0	0	0	0	0	1,04	0	0	0	0	0
0,740363	0	0	0	0	0	0,74	0,74	0	0	0	0
0,740363	0	0	0	0	0	0,74	1,48	0	0	0	0
1,041393	0	0	0	0	0	0	1,04	0	0	0	0
1,041393	0	0	0	0	0	0	1,04	0	0	0	0
1,041393	0	0	0	0	0	0	1,04	0	0	0	0
1,041393	0	0	0	0	0	0	1,04	0	0	0	0
1,041393	0	0	0	0	0	0	1,04	0	0	0	0
1,041393	0	0	0	0	0	0	1,04	0	0	0	0
1,041393	0	0	0	0	0	0	1,04	0	0	0	0
1,041393	0	0	0	0	0	0	0	1,04	0	0	0
1,041393	0	0	0	0	0	0	0	1,04	0	0	0
1,041393	0	0	0	0	0	0	0	1,04	0	0	0
1,041393	0	0	0	0	0	0	0	0	1,041	0	0
1,041393	0	0	0	0	0	0	0	0	1,041	0	0
1,041393	0	0	0	0	0	0	0	0	1,041	0	0
0,740363	0	0	0	0	0	0	0	0	0,74	0,74	0
1,041393	0	0	0	0	0	0	0	0	0	2,08	0

1,041393	0	0	0	0	0	0	0	0	0	1,04	0
1,041393	0	0	0	0	0	0	0	0	0	1,04	0
1,041393	0	0	0	0	0	0	0	0	0	1,04	0
1,041393	0	0	0	0	0	0	0	0	0	0	1,04
1,041393	0	0	0	0	0	0	0	0	0	0	1,04
1,041393	0	0	0	0	0	0	0	0	0	0	1,04
1,041393	0	0	0	0	0	0	0	0	0	0	1,04
skoring	9,12	16,8	4,79	4,82	12,1	7,21	9,51	3,21	7,051	6,03	4,91

Step 9. Based on the calculation of the total weight of sentences (D), 40% will be of the total sentence weight (D) which has the highest value.

Step 10. From the steps - these steps produce a collection of sentences (D) that contain important parts of the English-language document.

Then the results is:

1. The House's secretariat general said the legislative body had coordinated with the Saudi kingdom's protocol affairs office to find out all the things King Salman might need during his visit.
2. Damayanti, the secretariat general's deputy chairman, said physical preparations to welcome King Salman would take place over two weeks.
3. Built in 1968, certain parts of the Nusantara Building are unfriendly to the elderly and people with disabilities.
4. The House of Representatives is working to ensure that everything is well-organized for the visit of King Salman bin Abdulaziz al-Saud of Saudia Arabia.

RESULT

Here is survey for twenty four student in computer sciences, in 2020 years, on April 2020. The URL for survey is https://docs.google.com/forms/d/1Zj_p-hvLCfANQ7A0iufir89Y9NVaWoObS4zVbVeC80I/edit#responses, result for the survey, shown in figure 2 to figure 7

1. Have you ever read news in English? shown in figure 3 [ever, not ever]
2. Can you fully understand the content of news in English? Shown in figure 4 [full understand, alittle, most understand]
3. What do you think about reading news in english? Shown in figure 5 [dificult, most dificult, easy]
4. Have you ever summarized the news in English? Shown in figure 6 [ever, not ever]
5. How do you rate summarizing activities in English? Shown in figure 7 [dificult, easy, most sificult]
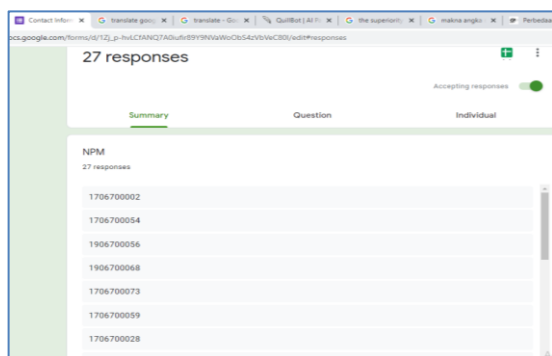
Figure 2 Twenty Seven Responden In Survey For Summary Activity

Figure 3 shows that 96 percent had read English language news,11.1 percent fully understood, 1 percent did not understand at all and 85.2 percent could fully understand (figure 4). according to 59.3 percent of news reading in English was not too difficult, 22.2 percent said it was difficult and 18.5 percent thought it was just the case (Figure 5). according to figure 6 that 77.8 percent of students have summarized a document in English and 22.2 percent said they had never. according to Figure 7 that 66.7 percent said summarizing English documents was difficult, 29.6 percent stated it was easy and 1 percent stated it was very difficult as in Figure 7.
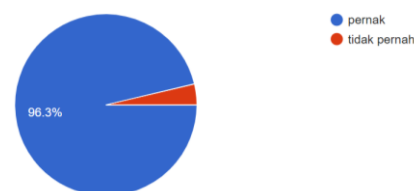
Figure 3 Ever Read News Media In English
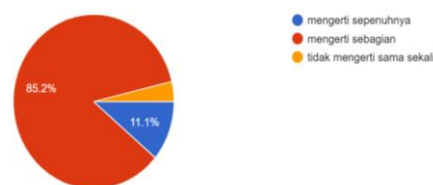
Figure 4 Experiences Made Summary fron English News Media
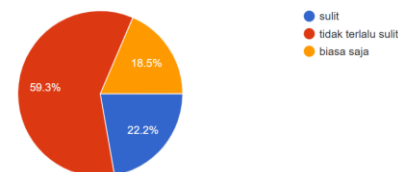
Figure 5 Perception Of Level In English News Media
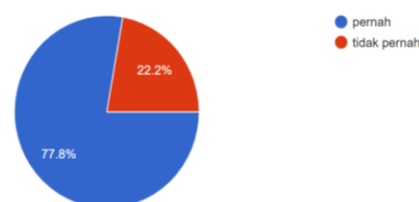
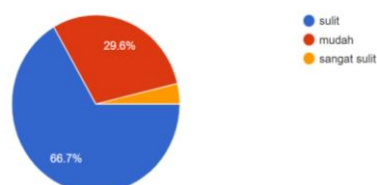Figure 6 Experiences Ever Read English News Media

Figure 7 Statistic Of Perception Computer Sciences Students To Summary Activity

Shown in figure 8 and figure 9, survey for text mining subject in STMIK Subang, there are fourty seventh students. Responden in survey was students that took text mining subject in last semester. The URL address is https://docs.google.com/forms/d/e/1FAIpQLSd5c6TgOv4C8950LVVPtuqLhiS9JGBd_tiz_vjLKhFnAFqxw/viewform .There are two questions as follows:

1. did you get enough description about the system of text summarizing in the text mining course? the option [not enough, enough]
2. Should lecturers demonstrate the application of the text summarization system in the beginning of the discussion? the option [interesting, not interesting, most interesting]

After the implementation of text mining courses 76.6 percent of students stated that they understood the contents of the course quite well, 23.4 percent said that it was not enough, as shown in Figure 8.

apakah anda mendapatkan gambaran yang cukup mengenai sistem peringkas teks dalam kajian mata kuliah teks mining?
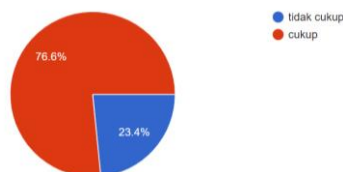
47 responses

- tidak cukup
- cukup

76.6%

23.4%

Figure 8 Student's Perception of Learning Text Mining Satisfactory

perlukan dosen mendemokan terlebih dahulu aplikasi sistem peringkas teks saat awal pembahasan?

47 responses

- perlu
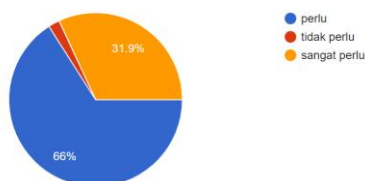- tidak perlu
- sangat perlu

31.9%

66%

Figure 9 Result Of Survey Demonstration Of Text Summarizer

In summary proses in this research, Precision is the level of accuracy between the information requested by the user and the answers given by the system, a summary of the English teacher's done with a summary by the system. While recall is the success rate of the system in finding back information.

The result from the summarizer system in this research is same as the result in material and method above, as follows:

1. The House's secretariat general said the legislative body had coordinated with the Saudi kingdom's protocol affairs office to find out all the things King Salman might need during his visit.

2. Damayanti, the secretariat general's deputy chairman, said physical preparations to welcome King Salman would take place over two weeks.

3. Built in 1968, certain parts of the Nusantara Building are unfriendly to the elderly and people with disabilities.

4. The House of Representatives is working to ensure that everything is well-organized for the visit of King Salman bin Abdulaziz al-Saud of Saudia Arabia.

For example evaluation of the English news Text Summarizing System using f-measure for the first responden on the table 3, as follows:

$$f - measure = \frac{2 \ x \ 1 \ x \ 0,75}{1+0,75} = 0,857143$$

On so on for another repondens

On table 4, For example evaluation of the English news Text Summarizing System using f-measure for tools4noobs, as follows:

$$f - measure = \frac{2 \ x \ 1 \ x \ 0,75}{1+0,75} = 0,857143$$

For example evaluation of the English news Text Summarizing System using f-measure for freesummarizer, as follows:

$$f - measure = \frac{2 \ x \ 1 \ x \ 0,5}{1+0,5} = 0,666667$$

And so on for another nines on line summarizer

Table 3. Testing by Twenty English Teachers

Number of Responden	Summariazer (P)	summary by teacher(B)	similarity (P∩Q)	Recall	precission	f-measure
1	4	3	3	0,75	1	0,857143
2	4	2	2	0,5	1	0,666667
3	4	2	2	0,5	1	0,666667
4	4	3	3	0,75	1	0,857143
5	4	4	4	1	1	1
6	4	2	2	0,5	1	0,666667
7	4	3	3	0,75	1	0,857143
8	4	3	3	0,75	1	0,857143
9	4	1	1	0,25	1	0,4
10	4	3	3	0,75	1	0,857143
11	4	3	3	0,75	1	0,857143
12	4	2	2	0,5	1	0,666667
13	4	4	4	1	1	1
14	4	3	3	0,75	1	0,857143
15	4	3	3	0,75	1	0,857143
16	4	2	2	0,5	1	0,666667
17	4	3	3	0,75	1	0,857143
18	4	4	4	1	1	1
19	4	2	2	0,5	1	0,666667
20	4	3	3	0,75	1	0,857143
total	80	55	55	0,6875	1	0,814815

Table 4. Summary by Online Summarizer

summerizer online	Summariazer (P)	by online summarizer(Q)	similarity (P∩Q)	recall	precission	f-measure
tools4noobs	4	3	3	0,75	1	0,857143
Resoomer	4	3	3	0,75	1	0,857143
Textcomparator	4	3	3	0,75	1	0,857143
SMMRY	4	3	3	0,75	1	0,857143
freesummarizer	4	2	2	0,5	1	0,666667
Summarygenerator	4	2	2	0,5	1	0,666667
simplish	4	3	3	0,75	1	0,857143
Summarizethis	4	2	2	0,5	1	0,666667
lakhasly	4	4	4	1	1	1
Autosummarixer	4	2	2	0,5	1	0,666667
Jumlah	40	27	27	0,675	1	0,80597

CONCLUSION

Based on the results obtained in this study, the conclusions obtained include:

- The English class in this study showed that summarizing 66,7 percent said it was difficult, so that additional media was needed in the form of a text summarizing system
- The method of demonstrating the English language text summarizing system is very preferably delivered at the beginning of the discussion of the text summarizing system in the text mining course as survey shown 66 percent is interesting, 31,9 percent is most interesting,
- The text summarization system has the feasibility of being used in learning activities as one of the media with a precision value of 1, a recall value of 0,6875. and an f-measure value of 0,814815 after testing by as many as twenty English teachers
- The text summarizing system has the feasibility of being used in text mining learning activities as one of the media in the demonstration method by obtaining a precision value of 1, a recall of 0,675 and an f-measure of 0,80597.

ACKNOWLEDGMENTS

REFERENCES

1. S. N. Maili and W. Hestiningsih, "Masalah-Masalah Pembelajaran Bahasa Inggris Pada Sekolah Dasar," *Media Penelit. Pendidik. J. Penelit. dalam Bid. Pendidik. dan Pengajaran*, vol. 11, no. 1, pp. 54–62, 2017.
2. S. Chelsky, "Kendala guru bahasa Inggris dalam penerapan k13 dan scientific approach di Indonesia," *Kompasiana.com*, 2019.
3. S. HENDRIANI, "Permasalahan Guru Bahasa Inggris Dalam Mengimplementasikan Kurikulum 2013 Di Slta Kabupaten Tanah Datar," *Ta'dib*, vol. 18, no. 1, p. 67, 2016.
4. F. Megawati, "Kesulitan Mahasiswa dalam Mencapai Pembelajaran Bahasa Inggris Secara Efektif," *Pedagog. J. Pendidik.*, vol. 5, no. 2, p. 147, 2016.
5. R. Ferreira-Mello, M. André, A. Pinheiro, E. Costa, and C. Romero, "Text mining in education," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 6, 2019.
6. F. Barbieri, F. Ronzano, and H. Saggion, "Summarization and information extraction in your tablet," *Proces. Leng. Nat.*, vol. 55, no. June 2016, pp. 203–206, 2015.
7. H. Li, J. Zhu, C. Ma, J. Zhang, and C. Zong, "Read, Watch, Listen, and Summarize: Multi-Modal Summarization for Asynchronous Text, Image, Audio and Video," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 996–1009, 2019.
8. H. H. Hsu, N. F. Huang, S. C. Chen, C. A. Lee, and J. W. Tzeng, "Misconceptions mining and visualizations for Chinese-based MOOCs forum based on NLP," *2017 IEEE 2nd Int. Conf. Big Data Anal. ICBDA 2017*, pp. 634–639, 2017.
9. Shodhganga@INFLIBNET, "CHAPTER 3:EVALUATION TECHNIQUES," pp. 39–52.
10. Jason Brownlee, "How to Calculate Precision , Recall , and F-Measure for Imbalanced Classification," *Mach. Leaening Mastery*, pp. 1–12, 2020.
11. S. A. Babar and P. D. Patil, "Improving performance of text summarization," *Procedia Comput. Sci.*, vol. 46, no. December, pp. 354–363, 2015.