# AUTOMATIC SUMMARIZATION TEXT ON MULTI DOCUMENT USING TEXTRANK

Rivandi Restu Putra[1,]Kania Dewi Evita, S.Pd., M.Sc.[2]
[1.2] Informatics Engineering Program
Faculty of Engineering and Computer Science University Computer Indonesia
Jl. Dipatiukur No. Bandung 112-114
E-mail: rivandirestu@gmail.com

## ABSTRACT

Along with the development of technology, more and more technologies that facilitate human work. One is the automatic text summary that facilitate human understanding of a text without reading overall. Automatic summarization captions are categories of *single-document* and *multi-document.* Summarization of *multi-document* becomes the focus of this study.

Automatic summarization captions multi summary document is text on some documents that have the same theme. In previous research on automatic captions summary on multi documents using RVM was found that the results of *the f-measure* is still low so the need to use other methods to improve accuracy.

In previous studies onsummary *single-document* using *TextRank* obtained an accuracy of 95% and good in summarizing a text message, so that in this study will be measured accuracy *TextRank* on summary multi document. From the results of tests performed, showed an accuracy of 74%, *a recall* of 73%, *precission* of 74%, and *f-measure* by 73%. Therefore we can concludealgorithm *TextRank* has *f-measure* a goodfor use in automatic captions summary multi document.

**Keywords:** Automatic Text summary, summary Automatic Multi Text Document, *TextRank,*Accuracy

## 1. INTRODUCTION

summary automatic captions*(automatictext*summarization)is the manufacture of a shorter version of a text by utilizing applications that run on the computer. The results of this summary contains important points of the original text. Summary defined as a text resulting from one or more text, containing information from the original text and is no longer than half the original texts [1].

In the multi-document summary research conducted by Tendi Arifin was found that the results of an accuracy of 67.35%, a recall of 47.83%, precission at 35.48, and f-measure amounted to 40.74% [2]. Having obtained the f-measure, it is still not good RVM method of its performance so that other methods are needed to improve the accuracy of the multi summary document.

Research conducted by Yuzar Marsyah states that the summary resultsalgorithm *textrank* contains a sentence that is more relevant to the results of 95.56% [3].algorithms *TextRank* are also faster in issuing the summary results of the algorithm LexRank 0.0294 seconds compared with 0.4321 seconds [3].

Based on the above, this thesis implements the algorithm *TextRank* in summary automatic text on a document using the Indonesian multi

### 1.1 Objectives

Based on the problems that have been studied, the purpose of the present study is to implement algorithms *textrank* summaryin the manufacture of automatic multi-document text.

While the purpose of this study was to measure the accuracy of automatic captions summary multidocument using the *textrank*algorithm.

### 1.2 Algorithm *TextRank*

*TextRank* is one of the graph-based ranking models to be applied in textual graph. Graph-based ranking algorithm is a way to determine how important a vertex in the graph, based on information extracted from the overall graph structure recursively. In its application to the automatic text summary, *TextRank* extracting high beranking sentences to be included in the summary [3].

### 1.3 TextPreprocessing

Text Pre-Processingis a stage that is carried out to prepare the source text into data to be processed at a later stage. Enter the type required in this process is in the form of a text document, the text will be through several major stages, namely :. The stages in the *text*processing,among others.

1. *Splitting,*news text intact will be split into several sentences.
2. Tokenizing said, cutting the sentence into single words ( "") spaces as delimiters.

3. *Filtering*, removing punctuation and symbols.
4. *Case folding*, the process of change throughout the report text to lowercase.
5. *Stopword Removal*, removing common words that frequently appear. These words are omitted because it has no meaning.

# 2 RESEARCH CONTENT

## 2.1 Problem Analysis

Problem analysis is an overview of the issues raised in writing on the implementation of the methodresearch *TextRank* on automatic text summary Indonesian language. In the multi-document summary research conducted by Tendi Arifin was found that the results of an accuracy of 67.35%, a recall of 47.83%, precission at 35.48, and f-measure amounted to 40.74% [2]. Having obtained the f-measure, it is still not good RVM method of its performance so that the necessary alternative methods for measuring the accuracy of the multi summary document.

## 2.2 Analysis System to be set up

automatic text analysis summary system is built by having three main stages, namely:

1. The first stage preprocessing process which is the stage where the document to be used is cleaned from noise or from things that do not have influence in the summary. Preprocessing is done in this phase include: Splitting, tokenizing, filtering, *Case Folding* and *stopword removal.*

2. The second stage is the algorithm *Textrank.*At this stage of initializing the weights beginning of a sentence and performing calculations *Overlap Content Similarity* to calculate the weight of the relationship between sentences. After that gocalculation *TextRank* to obtain any final weight of each sentence that will each weight will be in the rankings to get the final result of the summary.
3. The third step is to do the calculation TF-IDF and Cosine Similarity to sort the results of a summary of some of the documents and combine them into one results summary

## 2.3 Text Analysis Preprocessing

Given a word document that will be analyzed by the algorithm *textrank.*Examples of documents are as follows:

Timnas Islandia melanjutkan cerita ajaib mereka tahun ini. Setelah mencapai perempat final di Piala Eropa 2016 timnas Islandia untuk pertama kalinya masuk putaran final Piala Dunia 2018 di Russia. Lolosnya Islandia semakin luar biasa karena mereka lolos langsung tanpa perlu melalui fase play-off Piala Dunia 2018. Itu setelah Gylfi Sigurdsson dan kawan-kawan memuncaki Grup I dengan 22 poin dari 10 partai. Islandia sukses menyingkirkan Turki dan Ukraina yang notabene lebih berpengalaman. Mereka memaksa Kroasia harus melalui babak play-off sebelum memastikan diri lolos ke Piala Dunia 2018

Sample documents will be processed further ketahapan.

*1. Splitting,*news text intact will be split into several sentences.

Table 1 Splitting

| NO | Kalimat |
|---|---|
| S1 | Timnas Islandia melanjutkan cerita ajaib mereka tahun ini. |
| S2 | Setelah mencapai perempat final di Piala Eropa 2016, timnas Islandia untuk pertama kalinya masuk putaran final Piala Dunia 2018 di Russia. |
| S3 | Lolosnya Islandia semakin luar biasa karena mereka lolos langsung tanpa perlu melalui fase play-off Piala Dunia 2018. |
| S4 | Itu setelah Gylfi Sigurdsson dan kawan-kawan memuncaki Grup I dengan 22 poin dari 10 partai. |
| S5 | Islandia sukses menyingkirkan Turki dan Ukraina yang notabene lebih berpengalaman. |
| S6 | Mereka memaksa Kroasia harus melalui babak play-off sebelum memastikan diri lolos ke Piala Dunia 2018 |

*2. tokenizing* said, cutting the sentence into single words with ( "") spaces as delimiters

Table 2 *tokenizing*

| No | Hasil Tokenizing | | |
|---|---|---|---|
| S1 | Timnas<br>islandia<br>melanjutkan | cerita<br>ajaib<br>mereka | tahun<br>ini. |
| S2 | Setelah<br>mencapai<br>perempat<br>final<br>di<br>Piala<br>Eropa | 2016,<br>timnas<br>islandia<br>untuk<br>pertama<br>kalinya<br>masuk | putaran<br>final<br>piala<br>dunia<br>2018<br>di<br>Russia. |
| S3 | Lolosnya<br>Islandia | lolos<br>langsung | Piala<br>Dunia |

| | | | |
|---|---|---|---|
| | semakin<br>luar<br>biasa<br>karena<br>mereka | tanpa<br>perlu<br>melalui<br>fase<br>play-off | 2018. |
| S4 | Itu<br>setelah<br>Gylfi<br>Sigurdsson<br>dan<br>kawan-kawan | memuncaki<br>Grup<br>I<br>dengan<br>22<br>Poin | dari<br>10<br>Partai. |
| S5 | Islandia<br>sukses<br>menyingkirkan<br>Turki<br>dan | Ukraina<br>yang<br>notabene<br>lebih<br>berpengalaman. | |
| S6 | Mereka<br>memaksa<br>Kroasia<br>harus<br>melalui<br>babak | play-off<br>sebelum<br>memastikan<br>diri<br>lolos<br>ke | Piala<br>Dunia<br>2018. |

*3. Filtering*, removing punctuation and symbols.

Table 3 *Filtering*

| No | Hasil Filtering | | |
|---|---|---|---|
| S1 | Timnas<br>islandia<br>melanjutkan | cerita<br>ajaib<br>mereka | tahun<br>ini |
| S2 | Setelah<br>mencapai<br>perempat<br>final<br>di<br>Piala<br>Eropa | 2016<br>timnas<br>islandia<br>untuk<br>pertama<br>kalinya<br>masuk | putaran<br>final<br>piala<br>dunia<br>2018<br>di<br>Russia |
| S3 | Lolosnya<br>Islandia<br>semakin<br>luar<br>biasa<br>karena<br>mereka | lolos<br>langsung<br>tanpa<br>perlu<br>melalui<br>fase<br>playoff | Piala<br>Dunia<br>2018 |
| S4 | Itu<br>setelah<br>Gylfi<br>Sigurdsson<br>dan<br>kawan-kawan | memuncaki<br>Grup<br>I<br>dengan<br>22<br>poin | dari<br>10<br>Partai |
| S5 | Islandia<br>sukses<br>menyingkirkan<br>Turki<br>dan | Ukraina<br>yang<br>notabene<br>lebih<br>berpengalaman. | |
| S6 | Mereka<br>memaksa<br>Kroasia<br>harus<br>melalui<br>babak | playoff<br>sebelum<br>memastikan<br>diri<br>lolos<br>ke | Piala<br>Dunia<br>2018 |

*4. Case folding*, the process of change throughout the report text to lowercase.

Table 4 *Case Folding*

| No | Hasil Case Folding | | |
|---|---|---|---|
| S1 | timnas<br>islandia<br>melanjutkan | cerita<br>ajaib<br>mereka | tahun<br>ini |
| S2 | setelah<br>mencapai<br>perempat<br>final<br>di<br>piala<br>eropa | 2016<br>timnas<br>islandia<br>untuk<br>pertama<br>kalinya<br>masuk | putaran<br>final<br>piala<br>dunia<br>2018<br>di<br>russia |
| S3 | lolosnya<br>islandia<br>semakin<br>luar<br>biasa<br>karena<br>mereka | kolos<br>kangsung<br>tanpa<br>perlu<br>melalui<br>fase<br>playoff | piala<br>dunia<br>2018 |
| S4 | itu<br>setelah<br>gylfi<br>sigurdsson<br>dan<br>kawan-kawan | memuncaki<br>grup<br>i<br>dengan<br>22<br>poin | dari<br>10<br>partai |
| S5 | islandia<br>sukses<br>menyingkirkan<br>turki<br>dan | ukraina<br>yang<br>notabene<br>lebih<br>berpengalaman | |
| S6 | mereka<br>memaksa<br>kroasia<br>harus<br>melalui<br>babak | playoff<br>sebelum<br>memastikan<br>diri<br>lolos<br>ke | piala<br>dunia<br>2018 |

*5. Stopword Removal*, removing common words that frequently appear. These words are omitted because it has no meaning.

Table 5 *Stopword Removal*

| *No* | *Hasil Stopword Removal* | | Panjang Kalimat |
|---|---|---|---|
| S1 | timnas<br>islandia<br>cerita | ajaib<br>tahun | 5 |
| S2 | mencapai<br>perempat<br>final<br>piala<br>eropa<br>2016<br>timnas<br>islandia | kalinya<br>masuk<br>putaran<br>final<br>piala<br>dunia<br>2018<br>russia | 16 |
| S3 | lolosnya<br>islandia<br>lolos | dunia<br>2018 | 9 |

| | | | |
|---|---|---|---|
| | langsung<br>fase<br>playoff<br>piala | | |
| S4 | gylfi<br>sigurdsson<br>kawan-kawan<br>memuncaki<br>grup<br>i | 22<br>poin<br>10<br>partai | 10 |
| S5 | islandia<br>sukses<br>menyingkirkan<br>turki<br>ukraina | notabene<br>berpengalaman | 7 |
| S6 | memaksa<br>kroasia<br>babak<br>playoff<br>memastikan<br>diri | lolos<br>piala<br>dunia<br>2018 | 10 |

### *2.4* algorithm *Textrank*

*TextRank* one graph-based ranking models to be applied in textual graph. Graph-based ranking algorithm is a way to determine how important a vertex in a graph, based on information extracted from the overall graph structure recursively. In its application to the automatic text summary, *TextRank* extracting high beranking sentences to be included in the summary.

In the early stages ofalgorithm *textrank* thisis calculating the similarity / similarity between segments or between sentences with overlapping content, from making the same sentence on each segment of the sentence. If there are two words together in one sentence, then still count 1. Here is the formula of content overlap similarity:

$$Similarity\ (\mathrm{Si}, \mathrm{Sj}) = \frac{\left|\{\mathrm{Wk} \mid \mathrm{Wk} \in \mathrm{Si}\ \&\ \mathrm{Wk} \in \mathrm{Sj}\}\right|}{\log(|Si|) + \log(|\mathrm{Sj}|)}$$

Then, apply the formula of similarity of content overlap at all kalimat. This is a sample calculation *similarity of content overlap* for thesentences 2 and 3

Sim (S2, S3) = 4 / (log 16 + log 9) = 1.8532

Table 6 Similarity Overlap Conten

| | S1 | S2 | S3 | S4 | S5 | S6 |
|---|---|---|---|---|---|---|
| S1 | - | 1.0509 | 0.6049 | 0 | 0.6476 | 0 |
| S2 | 1.0509 | - | 1.3899 | 0 | 0.488 | 0.474 |
| S3 | 0.6049 | 1.8532 | - | 0 | 0.5558 | 2.1536 |
| S4 | 0 | 0 | 0 | - | 0 | 0 |
| S5 | 0.6476 | 0.488 | 0.5558 | 0 | - | 0 |
| S6 | 0 | 0.4746 | 2.1536 | 0 | 0 | - |

After the pre-processing and similarity, at this stage the weight calculation between sentences that will be ranked to get summary results. To calculate the weight of the sentence with *textrank* diperukan formula:

$$\mathrm{WS}(\mathrm{S_i}) = (1 - \mathrm{d}) + \mathrm{d} * \sum_{\mathrm{Vj} \in \mathrm{Adj(Vi)}} \frac{\mathrm{W_{ji}}}{\sum_{\mathrm{Vk} \in \mathrm{Adj(Vj)}} \mathrm{W_{jk}}} \mathrm{WS}(\mathrm{S_j})$$

Where:

d is a dumping factor which is filled with 0.85, $W_{ji}$ is the similarity of the words that you want to count, $W_{jk}$is connected with the sentence $W_j$, $S_j$ is filled with 1.

This formula is applied to all sentences to get the weight of each sentence that will result in the ranking.

Table 7 Ranking Sentence

| No. | Weight |
|---|---|
| S3 | 1.6962 |
| S2 | 1.1877 |
| S1 | 0.8472 |
| S6 | 0.6577 |
| S5 | 0.6113 |
| S4 | 0:15 |

If extraction of the sentence carried out 25% of the original text, then there are two words that go into the summary results, namely sentence 2 and sentence 3. After obtaining the results summary single document, this process is carried out as many documents, and all results are combined summary.

### 2.5 TF-IDF

Having summarized the results obtained, the sentence will be entered into TF-IDF is a sentence that has been done preprocesing process. After that, TF-IDF to get the weight of the word of the summary results. All results summary merged into one, on this theme there are 7 sentence summary results. After the counting occurrences of the word in every sentence

For example, the word "team"

1. To fill tf column, calculate the occurrence of the word "timnas" in every sentence. The word "timnas" only appears in the first sentence, so that in the TF-1 and TF worth R1-R were given a value of 0 to
2. Further on df column, add up the appearance of the word "team" in every sentence summary results. Because the team only appears one time in the first sentence, then df is 1
3. Fill idf column by using equation 2.3 by

entering the value of n and df. Value n = 7 and df = 1, so idf word "team" = log (7/1) = 0.8451
4. Fill column 2, using equation 2.4 by entering a value tf (the no.1) and idf (process no. 3). Thus filling W-R1 = tf-S1.idf = 1x w 0.8451 = 0.8451 and next.

### 2.6 Cosine Similarity

After Getting value tf-idf and values needed by the cosine similarity subsequently calculate the cosine similarity by using the formula:

$$Sim(R1, Ri) = \frac{\Sigma(R1, Ri)}{\sqrt{R1^2}\sqrt{Ri^2}}$$

Example calculation for Similarity to R1 and R2 are as follows:

$$\frac{0.60864}{\sqrt{10\ 011}\sqrt{3.702388}} = 0.099970915$$

for the calculation of the document can then use the same formula. After that the results obtained from the similarity of each document:

Table 8 Cosine Similarity

| | R1 | R2 | R3 | R4 | R5 | R6 | R7 |
|---|---|---|---|---|---|---|---|
| R1 | 0 | 0.099 | 0 | 0 | 0.012 | 0.046 | 0 |
| R2 | 0.099 | 0 | 0.040 | 0 | 0.013 | 0.064 | 0 |
| R3 | 0 | 0.040 | 0 | 0.049 | 0.121 | 0.036 | 0.016 |
| R4 | 0 | 0 | 0.049 | 0 | 0 | 0 | 0 |
| R5 | 0.012 | 0.013 | 0.121 | 0 | 0 | 0.050 | 0.048 |
| R6 | 0.046 | 0.064 | 0.036 | 0 | 0.050 | 0 | 0.045 |
| R7 | 0 | 0 | 0.01 | 0 | 0.048 | 0.045 | 0 |
| TOTAL | 0.158 | 0.217 | 0.263 | 0.049 | 0.244 | 0.243 | 0.110 |

On the results of the above table can be seen the value of every document that exists, then sorted from large to small and multi diapatkan results summary document.
Here is a sort of summary results of the multi-document:

Table 9 Ordering Results Summary

| No. | Sentence | Value | Document |
|---|---|---|---|
| 1 | R3 | 0.263992 | 2 |
| 2 | R5 | 0.244686 | 3 |
| 3 | R6 | 0.243739 | 4 |
| 4 | R2 | 0.217598 | 1 |
| 5 | R1 | 0.158503 | 1 |
| 6 | R7 | 0.110609 | 4 |
| 7 | R4 | 0.049438 | 2 |

### 2.7 Testing Functionality

Testing Functionality aims to showcase the functionality of the system in order to see the actual result in text summary system in the points to be tested. Each point of these tests will be explained back on the test results. Here is a table of test functionality:

Table 10 Testing Functionality

| No | Name Process | Points Test | Test Type |
|---|---|---|---|
| 1 | Process Select Scene News | Select news theme | *Blackbox* |
| | | Display text which will be summarized | *Blackbox* |
| 2 | Preprocessing | Showing sentence | *Blackbox* |
| | | Showing token | *Blackbox* |
| 3 | Process TextRank | Showing process TextRank | *Blackbox* |
| | | Results process TextRank | *Blackbox* |
| 4 | process TF-IDF and Cosine Similarity | Showing the TF-IDF and Cosine Similarity | *Blackbox* |
| | | Displays the results of the TF-IDF and Cosine Similarity | *Blackbox* |
| | | Showing summary multi document | *Blackbox* |

### 2.8 testing accuracy

After the rank of each sentence in each document, testing accuracy with comparing the results summary with the summary results of the system manually. Testing accuracy is a measure that indicates the proximity of the analysis results with

actual results. At this accuracy testing, quality is determined by thesummary. *accuracy*Rated *accuracy* obtained through the following formula:

$$Accuracy = \frac{\sum Kalimat\ Sistem\ yang\ relevan\ dengan\ Hasil\ Manual}{\sum Kalimat\ Sistem\ yang\ muncul}$$

to the application of this research, the number of sentences relevant systems with the results of the manual is 57 and the number of sentences the system that emerged was 77 so:

$$Accuracy = \frac{57}{77} = 0{,}74$$

Persentage 0,74 * 100% = 74%

Thus, the results obtained accuracy was 74%

### 2.9 Testing theperformance

Performance Testing was conducted in order to determine the result of applying the method *TextRank* in generating sentencessummary results. In this test will be used 24 news documents of the summary produced by the system will then be compared with the summary results of the manual produced by an Indonesian who is a graduate of Bachelor of Literature Indonesia, given the case studies that are taken in this study is a text document in Indonesian language [5] ,
Performance testing was performed under way to evaluate the results summary with intrinsic evaluation methods. This evaluation measures the value recall, precission and f-measure. This value is obtained from the following formula:

$$\boldsymbol{recall} = \frac{hits}{hits + misses}$$

Where:
$hits$ = Number of words in a sentence successfully extracted a system in accordance with the results of human extracts
$misses$ = Number of words in the sentence are extracted humans but not in the extract system

$$Precission = \frac{hits}{hits + noice}$$

Where:
$hits$ = Number of words in a sentence successfully extracted a system in accordance with the results of human extracts
$noice$ = Number of words in the sentence are extracted system but not in humans extracted sentence.

$$\boldsymbol{f\text{-}measure} = \frac{2 * precission * recall}{precission + recall}$$

Where:
*precission* = the result of the calculation *precission*
*Recall* = the result of the calculation *recall*

this formula is calculated to 24 documents being tested. Then obtained the following results:

Table 11 Test Performance

| Task Name | Avg_Recall | Avg_Precision | Avg_F-Score |
|---|---|---|---|
| DOKUMEN19 | 0.52 | 0.975 | 0.67826 |
| DOKUMEN18 | 0.83582 | 0.83582 | 0.83582 |
| DOKUMEN17 | 1 | 1 | 1 |
| DOKUMEN16 | 1 | 1 | 1 |
| DOKUMEN15 | 0.68657 | 0.66667 | 0.67647 |
| DOKUMEN14 | 1 | 1 | 1 |
| DOKUMEN13 | 0.72414 | 0.91304 | 0.80769 |
| DOKUMEN24 | 0.70455 | 0.70455 | 0.70455 |
| DOKUMEN12 | 0.68 | 0.59649 | 0.63551 |
| DOKUMEN23 | 0.66667 | 0.70968 | 0.6875 |
| DOKUMEN11 | 0.47059 | 0.35294 | 0.40336 |
| DOKUMEN22 | 0.97917 | 0.97917 | 0.97917 |
| DOKUMEN10 | 0.46939 | 0.62162 | 0.53488 |
| DOKUMEN21 | 0.98 | 0.98 | 0.98 |
| DOKUMEN20 | 0.97959 | 0.97959 | 0.97959 |
| DOKUMEN4 | 0.01653 | 0.03333 | 0.0221 |
| DOKUMEN3 | 0.67925 | 0.62069 | 0.64865 |
| DOKUMEN2 | 0.575 | 0.60526 | 0.58974 |
| DOKUMEN1 | 1 | 1 | 1 |
| DOKUMEN8 | 0.62069 | 0.36 | 0.4557 |
| DOKUMEN7 | 0.98507 | 0.98507 | 0.98507 |
| DOKUMEN6 | 0.41176 | 0.38889 | 0.4 |
| DOKUMEN5 | 1 | 1 | 1 |
| DOKUMEN9 | 0.58491 | 0.49206 | 0.53448 |
| Average | 0.732071 | 0.741661 | 0.730773 |

### 2.10 Evaluation test cohesion

Evaluation cohesion aims to calculate the coherence between sentences. If a value close to 0 then the summary is not coherent sentence, and if the value is close to 1 then it is already a coherent sentence summary [6]. Here are the results of the evaluation calculation cohesion:

$$CoH = \frac{Log(C_s * 9 + 1)}{Log\ (M * 9 + 1)}\ Ns = \frac{(o) * (o-1)}{2}$$

$$C_s = \frac{\Sigma_{\forall Si,Sj \epsilon Summary}\ Similarity\ (Si,Sj)}{N_s}$$

$$M = maxSimilarity\ (Si,Sj), Si,Sj\ \leq N$$

In this study, the formula is calculated on 5 themes, and obtained the following results:

Table 12 Test Evaluation Cohesion

| Scene | Cohesion Values |
|---|---|
| **2018 World Cup Qualifying** | **0.102966662** |
| **Hazard Transfer** | **0.322349521** |
| **Dismissal Arsene Wenger** | **0.24898729** |
| **Ozil Contract** | **0.326778014** |
| **Transfer Carton Cole** | **0.339610696** |

conclusion of the testing that has been conducted shows that the functional testing system using black box method for automatic text summary on multi document developed now running properly.

After the extraction of 25% to get the summary, obtained several elections in different words between system and user, this is because the weight gained by the system based solely on the number of the same words between sentences, while a summary is necessary to understand the meaning of the sentence as aspects of the formation of the results summary. This is a drawback of the algorithm *TextRank* who could not get the meaning of a sentence. Results of the accuracy of methods *TextRank* obtained average value recall 73%, precission 74%, and f-measure 73% based on testing 24 text document language news Indonesia

## 3 CONCLUSION

Based on research on summary automatic captions in multiple documents using *TextRank*,found that the method *TextRank* this can be implemented in the case of multi-document text summary. Results of testing to prove the value of accuracy generated by the algorithm TextRank is an accuracy of 74%, a recall of 73%, precission of 74%, and f-measure by 73% with the amount of news as much as 24 to obtain 57 the number of the same sentence between systems with manual from a total of 77 sentences on all documents.

Based on the results of testing that has been done, then the suggestions are expected after doing analysis on the multi summary automatic text documents using *TextRank* is between the test results with the manual system,algorithm *TextRank* has a weakness to understand the meaning of the sentence. So we need another method that can understand the meaning of a sentence to get the better accuracy

**REFERENCES**


[1] N. Purwasih, "Sistem Peringkas Teks Otomatis untuk Dokumen Tunggal Berita Berbahasa Indonesia dengan Menggunakan Graph-based Summarization Algorithm dan Similarity," 2009

[2] T. Arifin, "Peringkasan Teks Otomatis Pada Multi Dokumen Menggunakan Relevan Vector Machine," 2017.

[3] Y. Marsyah, "Perbandingan Kinerja Algoritme TextRank dengan Algoritme LexRank Pada Peringkasan Dokumen Bahasa Indonesia," 2013.

[4] M. Muchammad, A. Zainal and K. Ririen, "Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance," maret 2012.

[5] N. P. Lestari, "Uji Recall And Precission Sistem Temu Kembali Informasi OPAC Perpustakaan ITS Surabaya," 2016

[6] R. S. W. Rizki Darmawan, "Hybrid Keyword Extraction Algorithm and Cosine Similarity for Improving Sentences Cohesion in Text Summarization," 2015.