TEXT SUMMARIZATION FOR INDONESIAN LANGUAGE USING AHP

Ramdan Chandra M[1], Nelly Indriani W., S.Si.,M.T.[2]

[1,2]Teknik Informatika – Universitas Komputer Indonesia
Jalan Dipatiukur 112-116 Bandung
E-mail : [1]ramdanandchandra@gmail.com, [2]nelly.indriani@email.unikom.ac.id

ABSTRAK

Text summarization is a process of filtering most of the important information to produce a summary for the user. On Indonesian text summarization are used 6 features which are sentence frequency, keyword, word similarity with the title, similarity between sentences, sentence position and sentence length.
Method that used on Indonesian text summarization is AHP method. AHP method is considered suitable for text summarization, because this method will compare every feature and criteria, where criteria inside this text summarization is sentence that counted based on feature extraction, so that obtained final weight. In AHP method, feature values are obtained based comparison scale that obtained from an expert.
System testing on this Indonesian text summarization using AHP are done by using Rouge-1 which is a statistical testing method and expert consistency test on giving weight value. On rouge-1 method are obtained accuracy values for recall 75.83%, precision 63.61% and F-measure 68.59%. For statistical feature testing are obtained that f1 and f3 features has lowest accuracy and from expert consistency test are obtained value CR < 0.1, which means obtained comparison scale from expert is consistent.

Keywords: Text Summarization, Summary, AHP Method, Sentence

1. PREFACE

Text summarization is a process of filtering most of the important information to produce a summary for the user. Automati text summarization is a representation for a document to a shorter, but still have information on it[1]. Text summarization commonly seperated on 2 kinds which is extraction and abstraction. Extraction summarization is basically only took sentences that has important information on it. While abstraction summarization is a summarization that can produce new sentences that describes the main things from the summarized documents [2].

Actually, Indonesian text summarization have been researched with a various kind of methods. Which are Support Vecotr Machine Method (SVM) with Maximal Marginal Importance (MMI) [3], Relevance Vector Method (RVM) [4]. Those methods are method that is used in machine learning, where those methods are very affected by training and testing data. Training datas are used to train the alogirthm to find the suitable model and testing datas are used to know the performance from used model. Based on those reasons, this research will try to use decision making Analytical Hiearchy Process (AHP) method to be used as text summarization method.

Analytical Hierarchy Process (AHP) method are decision making method that can help human's framework of thinking. According to Bourgeois, AHP method usually used to prioritize from various alternatives where those options are complex or multi criteria [5].

Research on text summarization using AHP is ever done previously to summarize Perisan language [6]. In Persian AHP has good accuracy. With those good accuracy, AHP will be used for Indonesian text summarization which has a different grammatical structure with Persian.

Based on the description above, on this research will be used method on decision making system which is AHP, where priority selection decision will use one of expert as reference on deciding which important features to be use.

2. METHOD

This research will use *Analytical Hierarchy Process* (AHP) method for Indonesian text summarization. This method is a decision

making method. In Indonesian text summarization system, the first input is a document on *Portable Document Format* (PDF) file. When the document file is entered into the system, the file will be converted from binary on the pdf into text that later can be processed on preprocessing phase. After it enter preprocessing phase, the document will be processed in feature extraction phase. After the features are extracted, values from every features will be processed with AHP to get the summary result.

On this research, text summarization has 5 main phase in it system architecture. Which are text convertion, preprocessing, feature extraction, AHP and summary compression. Figure 1 is the full picture of architecture that is going to be built.

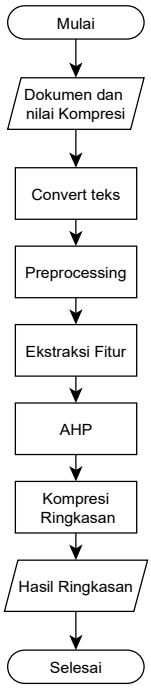


Figure 1 System Overview

2.1. Text Convert

On this phase will be the process of text convert which is an encoding on PDF which has binary into text. Those convert process is done by using Pdfparser library. After the data become text, the next phase is data sampling from the document that is going to be summarized. On figure 2 is process of data sampling of final project background as data that is going to be summarized.

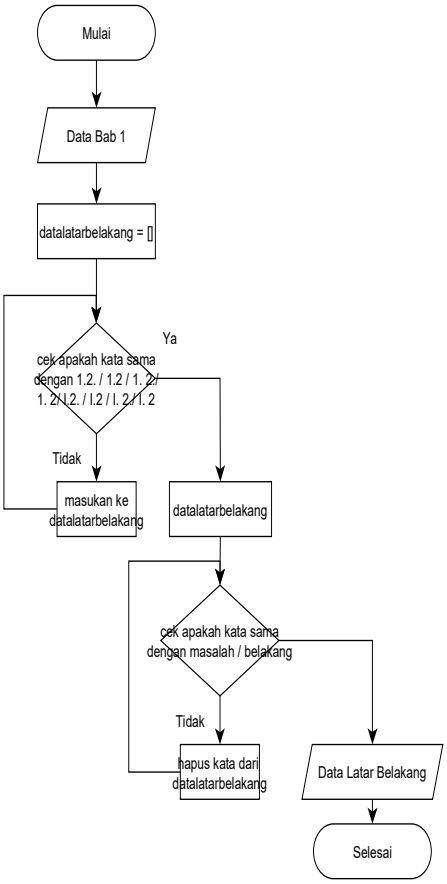


Figure 2 Background data retrieval process

2.2. *Preprocessing*

On this phase, after the entered data being convert, the data will be processed on preprocessing phase. Preprocessing itself has a function to throw away words or symbol that doesn't have any means on text summarization system. This process is consist of ssplitting the sentences, case folding, filtering, word tokenizing, stopword removal, and stemming [7]. On figure 3 is the steps of preprocessing process.

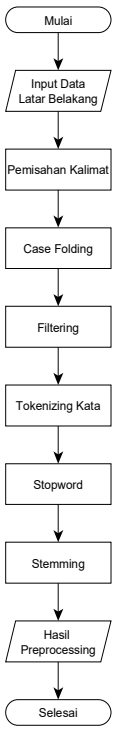


Figure 3 Preprocessing process

2.2.1 Sentence Separation

The first step on preprocessing is sentence separation. This process is used to seperate sentence in the document that is going to be summarized so that obtained a group of sentences. On this step, sentence separation are applied if on the document are found dot (.), question mark (?), and exclamation mark (!) [7]. For sentences that has it own words like (PT.), (CV.) etc, it will be transformed into (PT,), (CV,) so that sentence can be seperated.

2.2.2 *Case Folding*

After sentence separation, the next process is Case Folding. On this step every capital letters on the sentences will be transformed into small capital [7]. For example, the word "Kegiatan" will be transformetd into "kegiatan".

2.2.3 *Filtering*

Filtering process is used to throw away the punctuation on the text so that text only contains letters between 'a-z', numbers (0-9) and space (" ").

2.2.4 *Word Tokenizing*

This process is used to seperate word on textses. This process will seperate words on the sentences based on the space (" ") as seperator [7].

2.2.5 *Stopword removal*

Stopword removal is a process to remove words that doesn't have any meaning on it. To know what words that included in stopwords, are used stopword dictionary [8]. A sample of words that categorized as stopwords are "ke", "itu", "ini" etc.

2.2.6 *Stemming*

Stemming is a process to remove affix on word so that the word will become base word [9]. Method that will be used on this process is Indonesian Porter Stemmer [9].

2.3. Feature Extraction

After preprocessing the data will enter Feature Extraction phase. On this phase, sentence on document will be extracted based on the features. Used features are sentence frequency, keyword, word similarity with the title, sentences similarity, sentence location and sentence length.

2.3.1 Sentence Frequency

Sentence frequency feature can be extracted after preprocessing, where commonly the extraction is done by using tf-idf method on calculating word frequency [9]. Using formula (1).

$$wi = tfi * log \frac{\mathrm{N}}{\mathrm{n}} \qquad \textbf{(1)}$$

Where :

tfi : The number of certain word on the document
N : Total Document / Sentence
n : The number of document that contain searched word.

After (wi) values are calculated, the next step is feature calculation process for (F1) using formula (2) [6].

$$F1 = \frac{\sum_{i=1}^{k} Wi}{Max\ (Wi)} \qquad \textbf{(2)}$$

Where :

k : The number of words on the sentences

2.3.2 Keyword

On this phase keywords are obtained from words that appear several times, where words that categorized as keywords will be given 0.1 value and words that isn't found on the keywords will be given value 0 [7]. On keyword feature there is a condition where word on

keywords are not found the value will be 0. If there are sentences that has 0 value then the process cannot continue on AHP method calculation. So on keyword feature, every sentence will be summed up with 0.1 value.

2.3.3 Similarity of the word to the title

This feature will count words anywhere in a sentence similar to the one in the title. To calculate the resemblance of the word is used the formula (3).

$$\text{F3} = \frac{Kata\ dalam(i) \cap Kata\ dalam\ judul}{Kata\ dalam(i) \cup Kata\ dalam\ judul} + 0{,}1 \quad \textbf{(3)}$$

In this third feature there is a value of 0.1. A value of 0.1 on this feature is useful to prevent the value of feature 0 on the sentence.

For example, a title token of "giat ajar" with the first sentence is "giat, ajar, giat, individu". To calculate the value of features in the first sentence can be seen below:

$$F3 = \frac{(giat, ajar, giat, individu \cap giat, ajar)}{(giat, ajar, giat, individu \cup giat, ajar)}$$

$$F3 = \frac{(giat, ajar)}{(giat, ajar, giat, individu)}$$

$$F3 = \frac{2}{3} = 0{,}76667$$

in table 1 is a calculation of feature similarity of the title on the other sentence.

Tabel 1 Calculate Similarity of the word to the title

Sentence	F3
1	$F3 = \frac{2}{3} + 0{,}1 = 0{,}7667$
2	$F3 = \frac{2}{4} + 0{,}1 = 0{,}6$
3	$F3 = \frac{2}{7} = 0{,}385714$
4	$F3 = \frac{2}{7} + 0{,}1 = 0{,}385714$
5	$F3 = \frac{2}{5} + 0{,}1 = 0{,}5$
6	$F3 = \frac{2}{9} + 0{,}1 = 0{,}32222$

2.3.4 Similarity between sentences

This feature will calculate the resemblance of words in sentences with words in other sentences. This feature can be calculated by the formula (4).

$$F4 = \frac{Kata\ dalam(i) \cap Kata\ antar\ kalimat}{Kata\ dalam(i) \cup Kata\ antar\ kalimat} \quad (4)$$

For example, it is known the word in the first sentence is "giat, ajar, individu" and the word in the other sentence is "arah, ajar, bicara, didik, efektif, individu, jalan, giat, kembang, komponen, langsung, lihat, lintas, media, peserta, rencana, sekolah, strategi, tenaga". To get the value of feature similarity between sentences in the first sentence can be seen below:

$$F4 = \frac{(giat, ajar, individu \cap \begin{matrix} arah, ajar, bicara, didik, \\ efektif, individu, jalan, giat, \\ kembang, komponen, langsung, \\ lihat, lintas, media \\ peserta, rencana, sekolah, \\ strategi, tenaga \end{matrix})}{(giat, ajar, individu \cup \begin{matrix} arah, ajar, bicara, didik, \\ efektif, individu, jalan, giat, \\ kembang, komponen, langsung, \\ lihat, lintas, media \\ peserta, rencana, sekolah, \\ strategi, tenaga \end{matrix})}$$

$$F4 = \frac{(giat, ajar, individu)}{\left(\begin{matrix} arah, ajar, bicara, didik, \\ efektif, individu, jalan, giat, \\ kembang, komponen, langsung, \\ lihat, lintas, media \\ peserta, rencana, sekolah, \\ strategi, tenaga \end{matrix}\right)}$$

$$F4 = \frac{3}{21} = 0{,}142857$$

In table 2 is a calculation of feature similarity between sentences in other sentences.

Tabel 2 Calculation of Similarity between Sentences

Sentence	F4
S1	$F4 = \frac{3}{21} = 0{,}142857$
S2	$F4 = \frac{4}{21} = 0{,}190476$
S3	$F4 = \frac{7}{21} = 0{,}333333$
S4	$F4 = \frac{7}{21} = 0{,}333333$
S5	$F4 = \frac{5}{21} = 0{,}2380952$
S6	$F4 = \frac{9}{21} = 0{,}4285714$

2.3.5 Sentence Position

In the sentence position feature is based on the assumption that the first sentence and the last sentence on is the most important part of a document. The value of a feature at a sentence position can be calculated using the formula (5).

$$F5 = Max\left(\frac{1}{i}, \frac{1}{n - i + 1}\right) \quad \textbf{(5)}$$

Where i : Sentence Position

n : Count Sentence

The process of calculating the sentence position feature can be seen in table 3.

Tabel 3 Calculation of sentence position feature

Sentence	F5
1	$F5 = Max\left(\frac{1}{1}, \frac{1}{6-1+1}\right) = 1$
2	$F5 = Max\left(\frac{1}{2}, \frac{1}{6-2+1}\right) = 0{,}5$
3	$F5 = Max\left(\frac{1}{3}, \frac{1}{6-3+1}\right) = 0{,}33$
4	$F5 = Max\left(\frac{1}{4}, \frac{1}{6-4+1}\right) = 0{,}33$
5	$F5 = Max\left(\frac{1}{5}, \frac{1}{6-5+1}\right) = 0{,}5$
6	$F5 = Max\left(\frac{1}{6}, \frac{1}{6-6+1}\right) = 1$

2.3.6 Sentence Length

This feature is based on the fact that the longer the sentence means the more important words in the sentence. This feature can be calculated using the formula (6).

$$F6 = \frac{Sentence\ Lenth\ (i)}{the\ longest\ sentence\ in\ the\ document} \quad (6)$$

Dimana i : Sentence Position

the process of calculating the length of sentences can be seen in table 4.

Tabel 4 calculation on sentence length feature

Sentence	F6
1	$F6 = \frac{4}{13} = 0{,}3077$
2	$F6 = \frac{4}{13} = 0{,}3077$
3	$F6 = \frac{9}{13} = 0{,}69231$
4	$F6 = \frac{7}{13} = 0{,}5385$
5	$F6 = \frac{5}{13} = 0{,}48462$
6	$F6 = \frac{13}{13} = 1$

2.4. Analytic Hierarchy Process

The first step in this method is to scale the feature scale, which results from the feature scale obtained from expert experts. After the feature comparison is obtained, the feature value is entered into the matrix. Table 5 is a feature comparison matrix based on the comparative scale given by the expert.

Tabel 5 Matrix of feature comparison scale

Feature	F1	F2	F3	F4	F5	F6
F1	1	0,619	0,236	0,260	0,382	0,25
F2	1,613	1	0,381	0,420	0,617	0,403
F3	4,226	2,619	1	1,100	1,617	1,056
F4	3,839	2,380	0,908	1	1,469	0,959
F5	2,613	1,6199	0,6183	0,680	1	0,653
F6	4	2,479	0,946	1,041	1,530721	1

After a change is entered into the matrix, the next step is to normalize. The function of this normaslisasi in order to be in the range between 0 - 1.

Table 6 is the result of normalization based on the matrix in table 5.

Tabel 6 Feature Weight Results

	F1	F2	F3	F4	F5	F6	W
F1	0,052	0,085	0,041	0,038	0,038	0,038	0,049
F2	0,263	0,426	0,291	0,465	0,465	0,465	0,396
F3	0,052	0,0609	0,0416	0,031	0,031	0,031	0,041
F4	0,21	0,142	0,208	0,155	0,155	0,155	0,171
F5	0,21	0,142	0,208	0,155	0,155	0,155	0,171
F6	0,21	0,142	0,208	0,155	0,155	0,155	0,171

After the feature weighted value is obtained, proceed with searching the sentence weights on each feature. in table 7 is a calculation method to find the sentence weight of each feature.

Tabel 7 Calculation Weight Sentence of each feature

Fto(1..6)	S1	S2	...	Sn	W
S1	1	α			W1
S2	1/ α	1			W2
...			1		...
Sn				1	Wn

Table 8 is a matrix calculation on the first feature weights.

Tabel 8 Sentence Weight Matrix in Features 1

F1	S1	S2	S3	S4	S5	S6
S1	1	0,619	0,236	0,260	0,382	0,25
S2	1,6131	1	0,3816	0,420	0,617	0,403
S3	4,2262	2,6199	1	1,100	1,617	1,056

S4	3,8394	2,3800	0,9084	1	1,469	0,959
S5	2,6131	1,6199	0,6183	0,680	1	0,653
S6	4	2,4796	0,9464	1,041	1,530	1
$\sum sj$	17,29	10,719	4,091	4,503	6,61	4,328

The value 1 in the first column and the first line is obtained from the quotient between the value of the first sentence and the value of the second sentence. Value 0.619 obtained from the value of the first sentence divided by the value of the second sentence and so on until the sixth and sixth rows. After the matrix results obtained, matrix normalization is done so that the weight value exists between the range 0 - 1.

Table 9 is the result of normalization and weighting of the first feature.

Tabel 9 Normalization matrix of first feature weights

F1	S1	S2	S3	S4	S5	S6
S1	0,057	0,057	0,057	0,057	0,057	0,057
S2	0,09	0,09	0,09	0,09	0,09	0,09
S3	0,24	0,24	0,24	0,24	0,24	0,24
S4	0,22	0,22	0,22	0,22	0,22	0,22
S5	0,15	0,15	0,15	0,15	0,15	0,15
S6	0,23	0,23	0,23	0,23	0,23	0,23

After the matrix weighing the next feature is calculated weight value. Table 10 is the weight of feature 1.

Tabel 10 Bobot pada Fitur 1

Kalimat	W(F1)
S1	0,057
S2	0,09
S3	0,24
S4	0,22
S5	0,15
S6	0,23

For calculations on other features similar to the calculations performed on feature 1. Then after we get the value of sentence weight of each feature, the next process is to find the final weight, to get the final weight value used formula (7).

$$W_{Si} = \sum_{j=1}^{6} W_{Fj} * W_{ij} \quad (7)$$

Table 11 is the final result of the weight of every sentence in the exemplified document.

Tabel 11 Results The final weights

Kalimat	Wsi
S1	0,149693
S2	0,118922
S3	0,185164
S4	0,164302
S5	0,12924
S6	0,252679

2.5. Compression

at this stage a sentence will be chosen to be taken as a summary sentence. The compression value used between 30-40% [6]. Sentence is selected based on the highest value of weight. In the exemplary document it will use 30% compression. Thus, the sentence to be selected for the summary is 2 sentences and the sentence selected according to table 11 is sentence 3 and sentence 6. And the selected summary sentence can be seen in table 12.

Tabel 12 Summary results selected

Sentence	Result
S3	Ketika kita membicarakan tentang kegiatan pembelajaran ini maka kita mungkin akan langsung mengarahkan pandangan kita pada kegiatan pembelajaran yang berlangsung di sekolah.
S6	Beberapa diantaranya adalah mengenai tenaga pendidik, para peserta didik, media pembelajaran, strategi pembelajaran, dan perencanaan kegiatan pembelajaran

3. DISCUSSION

Testing is useful for knowing the results of accuracy when using the AHP method on Indonesian text abstracts. In this test using 50 test data on the background of the final project. Below there are 3 tests that are used Accuracy testing using Rouge with n-gram = 1, testing feature statistics and consistency testing from experts.

3.1 Pengujian akurasi *Rouge*

In this test using n-gram. In this method will calculate the value of recall, precision, and F-measure. In table 13 is the result of testing using Rouge.

Tabel 13 Test results using Rouge

Document	Recall	Precision	F-measure
1	55,435	59,302	57,303
2	73,529	54,945	62,893
3	68	62,044	64,885
4	89,61	76,67	82,635
5	76,74	66	70,968
6	96,67	50,87	66,67
7	84,615	49,749	62,658
8	97,826	72	82,949

9	71,579	81,928	76,404
10	72,864	66,514	69,544
11	78,378	67,442	72,5
12	86,301	92,647	89,362
13	86,184	66,497	75,072
14	74,576	57,895	65,185
15	73,563	83,117	78,049
16	100	100	100
17	94,167	84,962	89,328
18	53,425	34,513	41,935
19	74,684	63,441	68,605
20	84,524	66,981	74,737
21	56,835	33,193	41,91
22	73,016	63,014	67,647
23	71,765	52,586	60,697
24	75	50,336	60,0241
25	62,617	51,538	56,54
26	65	54,545	59,316
27	77,647	48,175	59,459
28	88,235	76,923	82,192
29	67,826	65	66,383
30	69,065	55,491	61,538
31	100	100	100
32	100	100	100
33	74,528	54,483	62,948
34	70,543	56,875	62,976
35	77,174	52,985	62,832
36	83,516	51,351	63,598
37	71,429	50	58,824
38	54,688	34,314	42,169
39	60,274	44,898	51,462
40	70,37	64,045	67,059
41	58,667	52,381	55,346
42	58,621	50	53,968
43	64,912	39,785	49,333
44	76,923	82,353	79,545
45	68	72,34	70,103
46	67,568	51,02	58,14
47	79,675	63,226	70,504
48	100	100	100
49	60,976	65,789	63,291
50	93,836	86,164	89,836
Rata - rata	75,82752	63,60654	68,58644

3.2 Statistical Feature Tests

In this feature it is useful to know how the feature can affect the accuracy value generated by the system. The statistical feature test will take 5 samples from 50 documents with the smallest F-measure value, this serves to know which features make the document has little accuracy. Table 14 is a statistical feature test.

Tabel 14 Testing Results Statistics Features

Features tested	Recall	Precision	F-measure
1	52,803	32,5514	40,244
1,2	64,552	39,54	48,98234
1,2,3	62,052	38,2746	47,27714
1,2,3,4	67,1028	41,7874	51,46714
1,2,3,4,5	63,5298	38,589	47,9908
1,2,3,4,5,6	61,3048	38,8026	47,4224

Based on the test results in Table 4.8. When using a feature that is sentence frequency feature (f1) recall value 52.803%, precision 32.5114% and F-measure 40.244%. Furthermore, the keyword features (f2) increase the value of recall, precision and F-measure respectively by 64,552, 39.54, and 48.98324. then added with the value of feature similarity of the word to the title (f3), resulting in impairment of recall, precision and F-measure is 62.052, 38.2746, 47.27714. Thereafter, a feature of similarity between sentences (f4) was added, after which there was a significant increase in the recall, precision and F-measure values of 67.1028, 41.7874, 51.46714. then added with the fifth feature of the position of the sentence, in testing this feature there is a decline in the value of recall 63.5298, precision 38.589, and F-measure of 47.9908. In the final step, the 6th feature is added to the length of the sentence. In this feature the recall and F-measure values are decreasing and the precision value has increased slightly with the recall value 61.308, precision 38.8036 and F-measure 47.4224. In this test it can be concluded that the features that have the greatest impact value are on the addition of the keyword feature (f2) and the sentence similarity (f4) feature and the less relevant feature are the sentence frequency (f1) and the word similarity to the title (f3) because it has a smaller accuracy value than the others.

3.3 Expert Consistency Testing

At this stage will be testing expert consistency in providing the feature value, whether the value provided by the expert can be used or not. Expert consistency is said to be consistent, if CR value <0.1. To get the CR value based on the explanation that has been done in the previous chapter [5].

The initial stage is to calculate the value of λmax. To get the λmax value begins by

multiplying each value in the first column by the value of the first feature weights then summing the value in the second column with the second feature weighted value then sum and so on. So we get the value in table 15.

Tabel 15 Calculation of weight values

Feature	Wight Value
F1	0,298222
F2	2,471845
F3	0,249927
F4	1,04939
F5	1,04939
F6	1,04939

After the weight value is found, the next step divides the weight value by the weighted value of the feature in table 19. so that the results obtained in table 16.

Tabel 16 Value

Feature	Value
F1	6,046188
F2	6,239394
F3	6,039387
F4	6,135282
F5	6,135282
F6	6,135282

Then from table 16 add each column and divided by the number of existing features to obtain λmaks value of 6.121802. To calculate the value of Ci and CR can be seen below:

Dik λ_{maks} = 6,121802
n = 6
Ri = 1,24

Dit Ci, CR ?

Jawab

$$Ci = \frac{\lambda_{\text{maks}} - \text{n}}{n-1}$$

$$Ci = \frac{6{,}121802 - 6}{6-1}$$

$$Ci = 0{,}02436$$

$$CR = \frac{\text{Ci}}{Ri}$$

$$CR = \frac{0{,}02436}{1{,}24}$$

$$CR = 0{,}0196$$

Based on the above calculation, it is known that CR value <0.1, it means that expert's given feature value is consistent.

4. Conclusions and Suggestions

4.1 Conclusions

Based on the results of previous research such as problem determination, problem analysis and implementation and testing, it can be concluded that the accuracy of the AHP method for the text of the Indonesian language has accurate results, using Rouge1 yields an average of 75.83% recall, precision 63.61%, and F¬-measure of 68.59%. .

4.2 Suggestions

Based on the results of the analysis and conclusion, there are suggestions for further research that is, Extraction features that are less important when compared by experts replaced with the extraction of other features. One of the less important feature extractions in this study is the similarity of the word to the title. Thus, it is worth replacing with other feature extraction.

BIBLIOGRAPHY

[1] Suyanto, Ririn Dwi Agustin Nengsih, "PERINGKASAN TEKS OTOMATIS UNTUK DOKUMEN TUNGGAL BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA TERM FREQUENCY DAN LEAD," Sekolah Tinggi Teknologi Telkom, Bandung, 2006.

[2] B. Zamandan E. Winarko, "Analisis Fitur Kalimat untuk Peringkas Teks Otomatis pada," *Peringkasan Text*, vol. 5, Juli 2011.

[3] Novriana, "Implementasi Support Vector Machine (SVM) Dan Maximal Marginal Importance (MMI) Untuk Peringkasan Multi Dokumen Artikel Berbahasa Indonesia," Unikom, Bandung, 2017.

[4] Arief Budiman Eka Putra, "Implementasi Metode Relevance Vector Machine Dalam Peringkasan Teks Otomatis," Unikom, Bandung, 2017.

[5] R Bourgeois, "*Analytical Hierarchy Process*," an Overview UNCAPSA - UNESCAP., Bogor, 2005.

[6] Ram Gopal Raj, Hamid Haj Seyyed Javadi Seyyed Mohsen Tofighy, "AHP Techniques for Persian Text Summarizatio," *Malaysian Journal of Computer Science*, vol. 26, pp. 1-8, 2013.

[7] R. Feldman and J. Sanger, "*The Text Mining Handbook*," Cambridge, 2007.

[8] Pandu Adikara, Putra, *Kamus Kata Dasar dan Stoplist Bahasa Indonesia,* http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/, 21 Agustus 2017 21.00

[9] Fadillah Z Tala, "A Study of Stemming Efects on Information Retrieval in Bahasa Indonesia," Institute for Logic, Language and Computation Universite itvan Amsterdam The Netherlands, Amsterdam, 2003.

[10] Sari Irene Nanda, "Analisis dan Implementasi Peringkasan Teks Otomatis dengan Metode Maximal Marginal Importance (MMI)," Univeristas Telkom, Skripsi 2011.