COMPARISON OF METHODS MAXIMUM MARGINAL RELEVANCE AND TEXTRANK ALGORITHMS ON THE SEARCH DOCUMENTS

Jajang Kamaludin[1], Nelly Indriani W., S.Si., M.T.[2]
[1,2] Program Studi Teknik Informatika
Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia
Jl. Dipatiukur No. 112-114 Bandung
E-mail : zarrotk@gmail.com[1], indiwidi@gmail.com[2]

ABSTRACT

Search document an affective way to get information automatically without having to read a whole and choose the relevant. May who do research on the method maximum marginal relevance and textrank algorithm and generates high accuracy. To know better then the algorithm to do the comparison. The comparison is the process of comparing the two algorithms to figure out a better accuracy of the method then maximum marginal relevance and algorithm textrank. From the result of testing that has been done, the comparison of result obtained by the method of maximum marginal relevance algorithm with textrank through the comparison of accuracy, and stated that the two algorithms has an accuracy rate of 100%. Thus it can be concluded that the two algorithms has a high degree of accuracy and is very suitable for use in document search news.

Keywords: Documen Search, Comparison Algorithms, MMR, Textrank, Accuracy.

1. INTRODUCTION

Information Retrieval (IR) is a method to find the back of unstructured data stored in a set of documents, and then presents information on the subject is needed. The aim of the IR system is to meet the needs of information users to obtain all the documents relevant to the user's needs and at the same time get a minimum of irrelevant documents.

Users can find relevant information by reading the entire document that existed at the depository, storing relevant documents, removing irrelevant documents, and sort documents in accordance with its requirements. It is an IR system is perfect, but this solution is not practical and efficient. Because users do not have much time to read the entire document one by one of the many existing documents.

Maximum marginal relevance (MMR) is one method of extraction summary (extractive summary) that is used to summarize a single document or multiple documents by calculating the similarity (simlarity) between parts of the text. Most of the information search engine generates a list ranking the relevancy of a decline in the demand while the user is done MMR method to approach the query relevancy of the documents by combining the cosine similarity matrix to rank the sentences in response to a user (query). Muhammad mustachiri According to research the accuracy of information retrieval percentage yield 100% [1].

Textrank algorithm is one method in a single automated text peringkasan using a graph-based ranking. Peringkasan algorithm graph-based document that there is TextRank algorithm [3]. TextRank graph assumes the fully-connected and non-directional graph with units of text (such as a word or phrase) as the vertex and the similarity between the text of the unit as the edge. Textrank itself has advantages not need any training data that is highly effective in the search of document in any language with any kind of (Mihalcea and Tarau 2004). According to research in the journal Yuzar Marsyah textrank comparison algorithm performance with algorithms lexrank the Indonesian peringkasan textrank states that the algorithm has a high accuracy rate of 95%.

Based on the above background, the reason to compare methods of maximum marginal relevance algorithm textrank already mentioned advantages and disadvantages that need to be researched their effectiveness in search of documents from the two algorithms are.

Based on the problems that have been studied, the purpose of the present study was to analyze the performance of the maximum marginal relevance methods and algorithms textrank to search documents.

The aim of this study was to measure the accuracy of the search for documents by the maximum marginal relevance methods and algorithms textrank.

1.1 Method MMR

Maximum marginal relevance (MMR) is one method of extraction summary (extractive summary) that is used to summarize a single document or multiple documents by calculating the similarity (simlarity) between the texts [1].

In the document search method to do the segmentation process MMR document into sentences

and do the grouping according to the gender of the sentence. MMR is used by combining the matrix cosine similarity to summarize a sentence in response to a query provided by the user.

1.2 Algorithm Texrank

Textrank algorithm is one method in a single automated text peringkasan using a graph-based ranking. Peringkasan algorithm graph-based document that there is TextRank algorithm [3]. TextRank graph assumes the fully-connected and non-directional graph with units of text (such as a word or phrase) as the vertex and the similarity between the text of the unit as the edge.

1.3 Text Processing

Text preprocessing is a stage for processing text information which is the raw material into the words that have been prepared are calculated weights he said. The stages on text processing, among others.

1. tokenizing sentence, the text of the news in the form pragraft will be split into several sentences.

2. Case folding is the process of converting uppercase to lowercase and removed all the punctuation, numbers and symbols.

3. Filtering, drop the word "less important or word that has no meaning.

4. tokenizing said, cutting the sentence based on the delimiters which it is composed, namely a space ("").

5. Stemming, restore a word to form roots (root word) with certain rules, so that each word has the same representation.

6. Weighting TF / IDF, the emergence of a term in a document term frequency (tf) and the number of occurrences of the terms in the document collection inverse document frequency (idf).

   IDF = log (D / dfi) (Formula 1)

   D is the number of documents that contain the term (t) and dfi is the number of occurrences (frequency) D. The term of the algorithm used to calculate the weight (W) of each document to keywords (query), namely:
   Wd, t = TFD, t * IDFT (Formula 2)

7. The use of cosine similiarity is the follow-up of TF-IDF weighting process. From the results it sought similarity weighting daintara existing documents. But its application in the MMR is the similarity of sentences in a document. The equation can be described as follows:

$$\mathrm{COS}(0ij) = \frac{\sum_k \mathrm{dikdjk}}{\sqrt{\sum_k d_{ik}^2}\sqrt{\sum_k d_{jk}^2}} \text{ (Rumus 3)}$$

2.CONTENTS RESEARCH

2.1 Analysis of the Problem

MMR method and the method of extraction document textrank is very effective at finding the document. Problems in the document search is how to determine the accuracy of the search of relevant documents.

In the document extraction method done tokenizing MMR document into sentences and do the grouping according to the gender of the sentence. MMR is done with a combination of cosine similarity matrix to rank the sentences in accordance with the response provided by the user. In the study conducted by Muchammad mustaqhfiri accuracy produced by the method MMR by a percentage of 100% with the highest and lowest 50% by value [1].

In the extraction document is done by yuzar marsyah on kekakurasian research generated by teritngi percentage of 95.56%, so it is necessary to study the accuracy of the comparison method and the maximum marginal relevance to the search algorithm textrank documents.

2.2 Analisis Sistem Yang Akan Dibangun

An overview of the stages to be performed by the system that is in Figure 1.

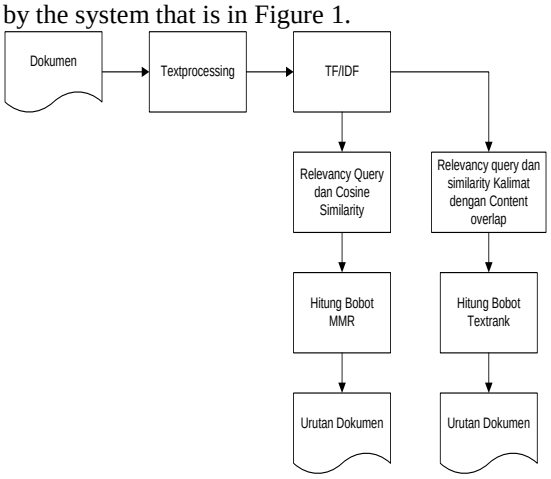


Figure 1 System Overview

2.3 Analisis Text Processing

Given a document d for the news to analysis by method MMR and textrank. Examples of the following documents:

Kepala Humas Badan Penyelenggara Jaminan Sosial Kesehatan Irfan Humaidi mengatakan iuran peserta mandiri atau pekerja bukan penerima upah (PBPU) mengalami kenaikan tersebut berlaku per 1 April 2016.

"Iuran peserta mandiri BPJS Kesehatan mengalami perubahan sebagaimana tertuang dalam peraturan

presiden yang baru diterbitkan. Masyarakat diharapkan dapat membacanya," katanya saat dihubungi Tempo, Sabtu, 12 Maret 2016.

Peraturan yang dimaksudkan Irfan adalah Peraturan Presiden Nomor 19 Tahun 2016 tentang Perubahan Kedua atas Peraturan Presiden Nomor 12 Tahun 2013 tentang Jaminan Kesehatan. Perpres itu sendiri diundang-undangkan pada 1 Maret lalu.

Dengan terbitnya perpres itu, besaran iuran kelas I yang semula Rp 59,500 menjadi Rp 80 ribu, iuran kelas II yang semula Rp 42,500 naik menjadi Rp 51 ribu, sedangkan iuran kelas III yang semula Rp 25,500 menjadi Rp 30 ribu.

Iuran peserta penerima bantuan iuran (PBI) serta penduduk yang didaftarkan pemerintah daerah juga mengalami kenaikan, dari sebelumnya Rp 19,225 menjadi Rp 23 ribu. Namun kenaikan iuran bagi peserta PBI tersebut sudah berlaku sejak 1 Januari lalu.

Irfan meminta penyedia fasilitas kesehatan yang bekerja sama dengan BPJS Kesehatan membaca perpres itu mengingat ada beberapa perubahan prosedur yang berhubungan dengan pelayanan kepada peserta Jaminan Kesehatan Nasional.

Examples of these documents will be in the next stage.

1. tokenizing sentences, sentences separated by a delimiter that connect such dots.

Table 1 tokenizing Sentence

NO	Kalimat
1	Kepala humas badan penyelenggara jaminan sosial kesehatan irfan humaidi mengatakan iuran peserta mandiri atau pekerja bukan penerima upah (pbpu) mengalami kenaikan tersebut berlaku per 1 april 2016
2	"Iuran peserta mandiri bpjs kesehatan mengalami perubahan sebagaimana tertuang dalam peraturan presiden yang baru diterbitkan
3	Masyarakat diharapkan dapat membacanya," katanya saat dihubungi tempo, sabtu, 12 maret 2016
4	Peraturan yang dimaksudkan irfan adalah peraturan presiden nomor 19 tahun 2016 tentang perubahan kedua atas peraturan presiden nomor 12 tahun 2013 tentang jaminan kesehatan
5	Perpres itu sendiri diundang-undangkan pada 1 maret lalu
6	Dengan terbitnya perpres itu, besaran iuran kelas i yang semula rp 59,500 menjadi rp 80 ribu, Iuran kelas ii yang semula rp 42,500 naik menjadi rp 51 ribu, Sedangkan iuran kelas iii yang semula rp 25,500 menjadi rp 30 ribu
7	Iuran peserta penerima bantuan iuran (pbi) serta penduduk yang didaftarkan pemerintah daerah juga mengalami kenaikan, dari sebelumnya rp 19,225 menjadi rp 23 ribu, namun kenaikan iuran bagi peserta pbi tersebut sudah berlaku sejak 1 januari lalu
8	Irfan meminta penyedia fasilitas kesehatan yang bekerja sama dengan bpjs kesehatan membaca perpres itu mengingat ada beberapa perubahan prosedur yang

	berhubungan dengan pelayanan kepada peserta jaminan kesehatan nasional

2. Case folding is the process of converting uppercase to lowercase and removed all the punctuation, numbers and symbols.

Table 2 Case Folding

NO	Kalimat
1	kepala humas badan penyelenggara jaminan sosial kesehatan irfan humaidi mengatakan iuran peserta mandiri atau pekerja bukan penerima upah pbpu mengalami kenaikan tersebut berlaku per april
2	iuran peserta mandiri bpjs kesehatan mengalami perubahan sebagaimana tertuang dalam peraturan presiden yang baru diterbitkan
3	masyarakat diharapkan dapat membacanya katanya saat dihubungi tempo sabtu maret

3. Filtering, drop the word "less important or word that has no meaning.Tokenizing kata, Pemotongan kalimat berdasarkan delimeter yang menyusunnya yaitu spasi (“ “).

Tabel 3 Tokenizing Kata

No	Segmen 5
1	perpres
2	sendiri
3	diundang
4	undangkan
5	maret
6	lalu

4. Stemming, restore a word to form roots (root word) with certain rules, so that each word has the same representation.

Table 4 Stemming

No	Segmen 5
1	perpres
2	sendiri
3	undang
4	undang
5	maret
6	lalu

5. Weighting TF / IDF, the emergence of a term in a document term frequency (tf) and the number of occurrences of the terms in the document collection inverse document frequency (idf).

Table 5 Calculation of TF / IDF

Term	Q	S1	S2	S3	S4	S5	S6	S7	S8	idf	Log(n/idf)
bpjs	1		1							1	TF=3 log(9/3)=0.477

Table 6 Calculation of Weights Terms

	Q	S1	S2	S3	S4	S5	S6	S7	S8
TF	1	0	1	0	0	0	0	0	1
IDF	0.477	0	0.477	0	0	0	0	0	0.477
W(T)=TF*IDF	0.477	0	0.477	0	0	0	0	0	0.477

After that, the overall weighted summation term in a particular segment, the segment S1 as follows.

$$\sum \mathrm{W(TF)} = 14.876$$

Information:

TF (Term Frequency): Number of Occurrences sentence

Table 7 Weight Relevancy Query

Q	S1	S2	S3	S4	S5	S6	S7	S8
0.732	14.876	5.03	6.729	7.785	4.469	10.43	11.065	15.256

Further multiplication of the query to each document (segmentation of the sentence), the S1 segment follows.

$$\Sigma(W(Q)*W(\mathrm{S1}))= 0.065$$

The second stage is to calculate the length of the sentence in any document including by squaring the query term weights each document, do Sum of squares then rooted so menghasil denominator in the equation.

$\sqrt{\sum(\mathrm{TF})^2}$ on segment 1 or sentence 1 is

$\sqrt{11.826}$ = 3.439

Overall results are obtained as follows.

Table 8 Long Sentence

Q	S1	S2	S3	S4	S5	S6	S7	S8
0.541	3.439	1.711	2.414	2.71	2.017	3.23	2.96	3.545

Then, apply the formula to calculate the cosine similarity a query similarity with other sentences. We take the example at S6, as follows.

$$\mathrm{COS}(Q, S6) = \frac{0.065}{0.0541 * 3.23} = 0.037$$

Overall results are obtained as follows.

Table 9 Value Results Similarity Between Sentences

	Q	S1	S2	S3	S4	S5	S6	S7	S8
Q	1.001	0.035	0.246	0	0	0	0.037	0.081	0.119
S1	0.035	1	0.133	0.051	0.07	0	0.006	0.132	0.11
S2	0.246	0.133	1	0	0.227	0	0.077	0.071	0.118
S3	0	0.051	0	1	0	0	0	0	0.1
S4	0	0.07	0.227	0	1	0	0	0	0.047
S5	0	0	0	0	0	1	0	0.071	0
S6	0.037	0.006	0.077	0	0	0	1	0.103	0.037
S7	0.081	0.132	0.071	0	0	0.071	0.103	1	0.012
S8	0.119	0.11	0.118	0.1	0.047	0	0.037	0.012	1

Maximum marginal relevance algorithm used to rank the sentences in response to a given query by the user. MMR calculations done by iterating that query to the overall similarity between the sentence and the sentence by sentence.

Users who want information space around the sample query, then it should set λ at a lower value. As for users who want to focus on strengthening the documents that make them more relevant, then it should set λ at a value closer to λ [1].

Sentences with the highest MMR value of each calculation iteration will be taken, which was chosen to serve as extraction document. Iteration will stop when the maximum MMR results equal to 0.

MMR

MMR = argmax[λ *Sim1(Si,Q) – (1- λ) * maxSim2(Si, S')]

Table 10 iterations MMR

iterasi	S1	S2	S3	S4	S5	S6	S7	S8
1	0.0315	0.2214	0	0	0	0.0333	0.0729	0.1071
2	0.3017	-	0	0.0227	0	0.0256	0.0658	0.0953
3	-	-	0	0	0.0227	0	0.0658	0.0953
4	-	-	0	0.0227	0	0.0256	0.0658	-
5	-	-	0	0.0227	0	0.0256	-	-
6	-	-	0	0.0227	0	-	-	-
7	-	-	-0.01	-	0	-	-	-

After that sort of result the overall iteration start early iterations until the final iteration to sort the highest yield to lowest.

Table 11 Results of Weight MMR

Iterasi ke	Kalimat	Bobot ArgMax MMR
1	S2	**0.3017**
2	S1	0.2214
3	S8	0.0953
4	S7	0.0658
5	S6	0.0256
6	S4	0.0227
7	S5	0

Results of the highest value weighting MMR is S2 with a weight of 0.3017 which will be used to display the document to the relevant query is invoked by the user.

2.4 Algorithm Textrank

Textrank algorithm is an algorithm of extraction sentences do terhadapa single document or multiple documents. Textrank a sentence processing algorithm to obtain the relevant documents in accordance responses given by the user.

To separate the contents of the document into sentences by using a single splitter form of punctuation such as a period ("."), Exclamation ("!"), A question mark ("?"), And newline. After separating the sentence by using a splitter, then calculate the similarity with overlapping content. Graph-based ranking method, vertex presentation sentences that are in the document while presenting edge similarity between two sentences that relate to each other so every single sentence obtained from the separation process sentences.

At this stage of calculating the weight of relevancy to the query that is fed by the user. After that, the overall weighted summation term in a particular segment, the segment Q as follows:

$$\sum W(TF) = 0.732$$

Tabel 12 Relevancy Query

Q	S1	S2	S3	S4	S5	S6	S7	S8
0.732	14.876	5.03	6.729	7.785	4.469	10.43	11.065	15.256

At this stage is to calculate the similarity / similarity between segments or between sentences with overlapping content, from making the same sentence on each segment of the sentence.

Further multiplication of the query to each document (sentence segmentation) on the S2 segment with 2.4 the following equation:$\Sigma(W(Q)*W(S2))$= 0.228

Table 13 Multiplication with Query

Q	S1	S2	S3	S4	S5	S6	S7	S8
0.293	0.065	0.228	0	0	0	0.065	0.13	0.228

The second stage is to calculate the length of each document including a query by the sum of the weights of each document term, then in the log. Log(SiSj) the second segment is 5.03= 0.702

Overall results are obtained as follows:

Table 14 Log (Si, Sj)

Q	S1	S2	S3	S4	S5	S6	S7	S8
-0.135	1.172	0.702	0.828	0.891	0.65	1.018	1.044	1.183

Then, apply the formula of similarity of content overlap 2.3, to calculate a similarity query with other sentences. Example on the S2, as follows:

Similarity (Q,S2)= 0.228/(-0.135+0.702)= 0.402

Table 15 Similarity Conten Overlap

	Q	S1	S2	S3	S4	S5	S6	S7	S8
Q	-1.085	0.063	0.402	0	0	0	0.074	0.143	0.218
S1	0.063	5.045	0.418	0.213	0.317	0	0.03	0.605	0.569
S2	0.402	0.418	2.085	0	0.662	0	0.248	0.205	0.379
S3	0	0.213	0	3.521	0	0	0	0	0.424
S4	0	0.317	0.662	0	4.121	0	0	0	0.219
S5	0	0	0	0	0	3.128	0	0.251	0
S6	0.074	0.03	0.248	0	0	0	5.123	0.479	0.194
S7	0.143	0.605	0.205	0	0	0.251	0.479	4.195	0.058
S8	0.218	0.569	0.379	0.424	0.219	0	0.194	0.058	5.311

Weighting process aims to get the final weight of a document. The more weight the end of a document, the more relevant the information in it.

$$WS(Si) = (1 - d) + d * \sum_{Sj \in Adj(Si)} \frac{Ws1si}{\sum_{SAdq(Sj)} Ws1q} WS(Sj)$$

After that sort iteration results overall from start S1 to S8 to sort the highest yield to lowest.

Table 16 Results Iteration Textrank

Iterasi	Bobot
S1	68.217
S7	25.085
S8	20.858
S2	6.187
S6	4.903
S4	4.427
S3	3.024
S5	1.642

From the above table can take the sentence to S1 as the highest value as a sample sentence that will be used to display the relevant documents.

The results of the analysis of MMR and Textrank is to collect all the weight of every document that is weighted to determine the relevant documents. Tests were conducted related to as many as 31. Query number of documents included in the experiments conducted is BPJS produce 5 documents with the highest weight to the query entered by the user. More can be seen in the table.

Table 17 Results Weight of MMR and Textrank

Dokumen	MMR	TEXTRANK
1	2.990	345.978
2	1.902	220.167
3	1.359	157.262
4	1.087	125.810
5	**0.543**	**62.905**

2.5 Testing

a. Plan testing to be performed on the application MMR and Textrank comparison algorithm is divided into 2 parts blackbox testing and testing of confusion matrix. Blackbox testing focuses on the commands executed by the conditions in the software and testing is done by the application builder. Testing confusion matrix focused on the results processed by the application to measure the accuracy of the results of the application.

b. Pengujian blackbox

Based on the results of black box testing (functional) to the test case above, it can be concluded that the development of applications comparison algorithm MMR and Textrank there are no errors. In the process of the application built functionally release the results in line with expectations.

Tabel 18 Hasil Pengujian Blackbox

Kasus dan Hasil Uji (Data Benar)			
Data Masukan	**Hasil yang diharapkan**	**Pengamatan**	**Kesimpulan**
Query: BPJS	Memproses dokumen yang terkait dengan query yang dimasukan	Sistem menampilkan dokumen yang terkait dengan query sesuai yang diberikan oleh user	[√] Berhasil [ ] Tidak Berhasil
Kasus dan Hasil Uji (Data Salah)			
Query: csadasd	Sistem menampilkan pesan "query tidak dapat Ditemukan".	Tidak dapat menampilkan Dokumen	[√] Berhasil [ ] Tidak Berhasil

c. Pengujian MMR

MMR testing is carried out by entering a query first by example query "BPJS" against 35 news documents.

Table 19 Testing MMR

No	Judul Dokumen Berita	Nilai
1	Tunggakan iuran bpjs kesehatan mandiri mencapai rp 18miliar	TRUE
2	Bpjs kesehatan akan langsung nonaktifkan penunggak iuran	TRUE
3	Bpjs TK tanggung biaya pengobatan korban kerja normalisasi kali ciliwung	TRUE
4	Peserta mandiri bpjs kesehatan kini bisa bayar iuran kolektif	TRUE
5	iuran bpjs naik	TRUE
6	Prabowo-sohibul iman bahas demonstrasi 4 november	FALSE
7	Antsari didorong bongkar dugaan kaitan SBY dengan kasusnya	FALSE
8	TNI siap ikut amankan demo 4 november	FALSE
9	Pengamanan wilayah RI, kebutuhan ideal brimob 3.000 personel	FALSE
10	Polisi diminta kedepankan sikap humanis tangani demo 4 november	FALSE
11	Komentar politikus pdip terkait panglima tni	FALSE
12	Penjelasan mabes polri soal tindakan tembak di tempat demonstran	FALSE
13	Jokowi minta aparat siaga kawal aksi 4 november	FALSE
14	Protes penistaan agama, dpr minta pejabat hati-hati berucap	FALSE
15	Surat siaga I bocor picu kecemasan masyarakat	FALSE

16	Pemerintah diminta terbitkan peraturan kartu handphone khusus anak	FALSE
17	Wirausahaan hadapi sejumlah kendala, ini saran HT	FALSE
18	Cair klaim asuransi 3.458 TKI korban PHK binladin group	FALSE
19	Mudahkan calon jamaah umrah, amphuri go diapresiasi	FALSE
20	Harapan jokowi kepada ullama	FALSE
21	Kabar BJ Habibie meninggal tidak benar	FALSE
22	Indonesia harus cetak entrepreneur produktif secara masif	FALSE
23	Kemendikbud diminta tak paksakan sekolah terapkan full day school	FALSE
24	Ombudsman nilai pemilihan rektor ptn rawan permainan uang	FALSE
25	37 perguruan tinggi adu jago karya inovatif	FALSE
26	Pemerintah nilai pendidikan vokalis bisa tekan pengangguran	FALSE
27	16 siswa sd akan dipilih jadi peserta ASMOPS 2016	FALSE
28	UI-Unbraw gelar konfrensi Internasional ilmu komputer	FALSE
29	Seminar nasional negara terlibat pembantaian massal 1965-1966	FALSE
30	Oposisi iran khamenei yang perintahkan houthi serang makkah	FALSE
31	TNI AU-marinir AS gelar latihan bersama di manado	FALSE
32	As sempat kerahkan senjata aneh pencegah kiamat nuklir korut	FALSE
33	FBI usik lagi skandal E-mail, Hillary dikalahkan Trump	FALSE
34	Politisi iran bantah sebut perang dengan israel berarti bunuh diri	FALSE
35	Gempa bumi 6,1 SR guncang manado, warga berhamburan	FALSE

MMR testing in the table with the query "BPJS" produces the correct search as much as 5 news documents, 30 documents the news is not true, and the absence of missing results in the search document.

Table 20 Results of Confusion Matrix

	Nilai Sebenarnya	
	TRUE	FALSE
TRUE	(TP) 5	(FP) 0
FALSE	(FN) 0	(TN) 30

Precision $= \frac{TP}{TP+FP}$

$= \frac{5}{5+0} = 100\%$

Recall $= \frac{TP}{TP+FN}$

$= \frac{5}{5+0} = 100\%$

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$

$= \frac{5+30}{5+30+0+0} = 100\%$

d. Testing Textrank

Table testing textrank confusion matrix, stating the correct search results there are 5 documents out of a total of 35 news news documents, as well as the results do not match, there are 30 news.

Table 21 MMR Testing Results

No	Judul Dokumen Berita	Nilai
1	Tunggakan iuran bpjs kesehatan mandiri mencapai rp 18miliar	TRUE
2	Bpjs kesehatan akan langsung nonaktifkan penunggak iuran	TRUE
3	Bpjs TK tanggung biaya pengobatan korban kerja normalisasi kali ciliwung	TRUE
4	Peserta mandiri bpjs kesehatan kini bisa bayar iuran kolektif	TRUE
5	iuran bpjs naik	TRUE
6	Prabowo-sohibul iman bahas demonstrasi 4 november	FALSE
7	Antsari didorong bongkar dugaan kaitan SBY dengan kasusnya	FALSE
8	TNI siap ikut amankan demo 4 november	FALSE
9	Pengamanan wilayah RI, kebutuhan ideal brimob 3.000 personel	FALSE
10	Polisi diminta kedepankan sikap humanis tangani demo 4 november	FALSE

11	Komentar politikus pdip terkait panglima tni	FALSE
12	Penjelasan mabes polri soal tindakan tembak di tempat demonstran	FALSE
13	Jokowi minta aparat siaga kawal aksi 4 november	FALSE
14	Protes penistaan agama, dpr minta pejabat hati-hati berucap	FALSE
15	Surat siaga I bocor picu kecemasan masyarakat	FALSE
16	Pemerintah diminta terbitkan peraturan kartu handphone khusus anak	FALSE
17	Wirausahaan hadapi sejumlah kendala, ini saran HT	FALSE
18	Cair klaim asuransi 3.458 TKI korban PHK binladin group	FALSE
19	Mudahkan calon jamaah umrah, amphuri go diapresiasi	FALSE
20	Harapan jokowi kepada ullama	FALSE
21	Kabar BJ Habibie meninggal tidak benar	FALSE
22	Indonesia harus cetak entrepreneur produktif secara masif	FALSE
23	Kemendikbud diminta tak paksakan sekolah terapkan full day school	FALSE
24	Ombudsman nilai pemilihan rektor ptn rawan permainan uang	FALSE
25	37 perguruan tinggi adu jago karya inovatif	FALSE
26	Pemerintah nilai pendidikan vokalis bisa tekan pengangguran	FALSE
27	16 siswa sd akan dipilih jadi peserta ASMOPS 2016	FALSE
28	UI-Unbraw gelar konfrensi Internasional ilmu komputer	FALSE
29	Seminar nasional negara terlibat pembantaian massal 1965-1966	FALSE
30	Oposisi iran khamenei yang perintahkan houthi serang makkah	FALSE
31	TNI AU-marinir AS gelar latihan bersama di manado	FALSE
32	As sempat kerahkan senjata aneh pencegah kiamat nuklir korut	FALSE
33	FBI usik lagi skandal E-mail, Hillary dikalahkan Trump	FALSE
34	Politisi iran bantah sebut perang dengan israel berarti bunuh diri	FALSE
35	Gempa bumi 6,1 SR guncang manado, warga berhamburan	FALSE

The test results with the query "BPJS" news documents as 35, produces the correct value by 5 documents and 30 documents were wrong.

Table 22 Results Matrix Confusin Textrank

	Nilai Sebenarnya	
	TRUE	FALSE
TRUE	(TP) 5	(FP) 0
FALSE	(FN) 0	(TN) 30

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{5}{5+0} = 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{5}{5+0} = 100\%$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{5+30}{5+30+0+0} = 100\%$$

Based on the results of testing that has been done can be concluded that the software is functioning as expected from the start as well as software that built have met the initial goals of development. Textrank algorithms tend to be slow in

processing documents compared with MMR algorithm.

In the test with the query "BPJS" produce convenient search as much as 5 news of total document and produce news 35 30 news that does not fit. The accuracy of the algorithms are very good average results of a document search has a 100% accuracy rate.

3 Covering

Based on research on the MMR and Textrank comparison algorithm, testers managed to determine the accuracy of which is generated by the two algorithms are. Results of the test proved the value of accuracy generated by the algorithm produces accuracy of 100% by the number of news documents as 35, and produces accuracy corresponding query as much as 5 news documents and 30 news that does not fit, the search time produced by MMR faster than textrank appropriate with the number of documents processed.

Based on the results of testing that has been done, then the suggestions are expected after doing an analysis and comparison algorithms and algorithms Textrank MMR.

1. Further development wearing a title generator, to ensure the contents of the document and the same title has relevance, because the processing of more documents to fill in the sentence.
2. It is recommended to test the document even more.
3.

DAFTAR PUSTAKA

[1] M. Mustaqhfiri, Z. Abidin And R. Kusumawati, "Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance," Matics, Vol. Iv, No. 4, Pp. 1-14, 2011.

[2] D. Nopiyanti And K. A. Sekarwati, "Aplikasi Pencarian Kata Dasar Dokumen Berbahasa Indonesia Dengan Metode Stemming Porter Menggunakan Php & Mysql," In Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen (Kommit 2014) , Depok, 2014.

[3] Yuzar Marsyah (2013). Perbandingan Kinerja Algoritma Textrank Dengan Algoritma Lexrank Pada Peringkasan Dokumen Bahasa Indonesia. [Online]. Diunduh pada tanggal 30 april 2016.