PERBANDINGAN METODE *MAXIMUM MARGINAL RELEVANCE* DAN ALGORITMA *TEXTRANK* PADA PENCARIAN DOKUMEN

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]

ABSTRAK

Pencarian dokumen merupakan cara efektif untuk mendapatkan informasi secara otomatis tanpa harus membaca keseluruhan dan memilih dokumen yang relevan. Banyak yang melakukan penelitian terhadap metode maximum marginal relevance dan algoritma textrank dan menghasilkan keakurasian yang tinggi . Untuk mengetahui algoritma yang lebih baik maka perlu dilakukan perbandingan. Perbandingan merupakan proses membandingkan kedua algoritma untuk mengetahui tingkat akurasi yang lebih baik dari metode *maximum marginal relevance* dan algoritma *textrank*. Dari hasil pengujian yang telah dilakukan, didapatkan hasil perbandingan metode maximum marginal relevance dengan algoritma textrank melalui perbandingan akurasi, dan menyatakan bahwa kedua algoritma memiliki tingkat akurasi sebanyak 100%. Sehingga dapat disimpulkan bahwa kedua algoritma memiliki tingkat akurasi yang tinggi dan sangat cocok untuk digunakan dalam pencarian dokumen berita.

Kata kunci : pencarian dokumen, perbandingan algoritma, *MMR*, *textrank*, akurasi

1. PENDAHULUAN

Information Retrieval (IR) merupakan suatu metode untuk menemukan kembali data tidak terstruktur yang tersimpan pada sekumpulan dokumen, kemudian menyediakan informasi mengenai subyek yang dibutuhkan. Tujuan dari sistem IR ini adalah memenuhi kebutuhan informasi pengguna dengan mendapatkan semua dokumen yang relevan dengan kebutuhan pengguna dan pada waktu yang sama mendapatkan sesedikit mungkin dokumen yang tak relevan.

Pengguna dapat menemukan informasi yang relevan dengan membaca seluruh dokumen yang ada pada tempat penyimpanannya, menyimpan dokumen-dokumen yang relevan, membuang dokumen yang tidak relevan, dan mengurutkan dokumen-dokumen yang sesuai dengan keperluannya. Hal tersebut merupakan sistem IR yang sempurna, tetapi solusi ini tidak praktis dan efisien. Dikarenakan pengguna tidak memiliki banyak waktu untuk membaca seluruh dokumen satu per satu dari sekian banyak dokumen yang ada.

Maximum marginal relevance (MMR) merupakan salah satu metode ekstraksi ringkasan (*extractive summary*) yang digunakan untuk meringkas dokumen tunggal atau multi dokumen dengan menghitung kesamaan (*simlarity*) antara bagian teks. Kebanyakan mesin pencarian informasi menghasilkan daftar perangkingan dari penurunan *relevancy* terhadap permintaan user sedangkan yang dilakukan metode MMR melakukan pendekatan terhadap relevancy query terhadap dokumen dengan mengkombinasikan matrik cosine similarity untuk merangking kalimat-kalimat sebagai tanggapan dari user (query). Menurut muhammad mustachiri dengan penelitiannya ke akuratan hasil persentase pencarian informasi 100%[1].

Algoritma *Textrank* adalah salah satu metode dalam peringkasan teks otomatis tunggal menggunakan pemeringkatan berbasis graf. Algoritma peringkasan dokumen berbasis graf yang ada adalah algoritma TextRank[3]. TextRank mengasumsikan graf yang *fully-connected* dan graf tidak berarah dengan unit teks (seperti kata atau kalimat) sebagai *vertex* dan *similarity* antar teks unit sebagai *edge*. Textrank sendiri memiliki kelebihan tidak perlu adanya data latih yang sangat efektif dalam pencarian dokumen dengan segala jenis bahasa apapun (Mihalcea dan Tarau 2004). Menurut penelitian Yuzar Marsyah dalam jurnal *perbandingan kinerja algoritma textrank dengan algoritma lexrank pada peringkasan bahasa indonesia* menyatakan bahwa algoritma textrank memiliki tingkat akurasi tinggi sebesar 95%.

Berdasarkan latar belakang di atas, alasan membandingkan metode maximum marginal relevance dengan algoritma textrank yang sudah disebutkan kelebihan dan kekurangannya sehingga perlu diteliti keefektifannya dalam pencarian dokumen dari kedua algoritma tersebut.

1.1 Maksud dan Tujuan

Berdasarkan permasalahan yang telah diteliti, maka maksud dari penelitian kali ini adalah melakukan analisis kinerja terhadap metode *maximum marginal relevance* dan algoritma *textrank* untuk pencarian dokumen.

Sedangkan tujuan penelitian ini adalah untuk mengukur keakurasian pada pencarian dokumen oleh metode *maximum marginal relevance* dan algoritma *textrank*.

1.2 Metode MMR

Maximum marginal relevance (MMR) merupakan salah satu metode ekstraksi ringkasan (*extractive summary*) yang digunakan untuk meringkas dokumen tunggal atau multi dokumen dengan menghitung kesamaan (*simlarity*) antara bagian teks[1].

Pada pencarian dokumen dengan metode MMR dilakukan proses segmentasi dokumen menjadi kalimat dan dilakukan pengelompokan sesuai dengan gender kalimat tersebut. MMR digunakan dengan mengkombinasikan matrik *cosine similarity* untuk merangkum kalimat sebagai tanggapan pada *query* yang diberikan oleh *user*.

1.3 Algoritma *Texrank*

Algoritma *Textrank* adalah salah satu metode dalam peringkasan teks otomatis tunggal menggunakan pemeringkatan berbasis graf. Algoritma peringkasan dokumen berbasis graf yang ada adalah algoritma TextRank[3]. TextRank mengasumsikan graf yang *fully-connected* dan graf tidak berarah dengan unit teks (seperti kata atau kalimat) sebagai *vertex* dan *similarity* antar teks unit sebagai *edge*.

1.4 Text Processing

Text preprocessing adalah suatu tahap untuk mengolah teks informasi yang merupakan bahan mentah menjadi kata-kata yang telah siap dihitung bobot katanya. Adapun tahapan pada *text processing*, diantaranya.

1. *Tokenizing* Kalimat, teks berita yang berupa pragraft akan dipisah menjadi beberapa kalimat-kalimat.
2. *Case folding* merupakan Proses mengubah huruf kapital menjadi huruf kecil dan membuang semua tanda baca, angka dan simbol.
3. *Filtering*, membuang kata" yang kurang penting atau kata yang tidak mempunyai makna.
4. Tokenizing kata, Pemotongan kalimat berdasarkan delimeter yang menyusunnya yaitu spasi (" ").
5. *Stemming,* mengembalikan suatu kata ke bentuk akarnya (*root word*) dengan aturan-aturan tertentu, sehingga setiap kata memiliki representasi yang sama.
6. Pembobotan TF/IDF, kemunculan suatu *term* dalam sebuah dokumen *term frequency* (*tf*) dan jumlah kemunculan *term* dalam koleksi dokumen *inverse document frequency* (*idf*).

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

   D adalah jumlah dokumen yang berisi *term* (*t*) dan *dfi* adalah jumlah kemunculan (frekuensi) *term* terhadap D. Adapun algoritma yang digunakan untuk menghitung bobot (W) masing-masing dokumen terhadap kata kunci (*query*), yaitu:

   Wd,t = tfd,t * IDFt (Rumus 2)
7. Penggunaan *cosine similiarity* adalah tindak lanjut dari proses pembobotan TF-IDF. Dari hasil pembobotan maka dicarilah kemiripan daintara dokumen yang ada. Namun penerapannya dalam *MMR* adalah kemiripan kalimat-kalimat dalam sebuah dokumen. Persamaannya dapat digambarkan sebagai berikut:

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

2 ISI PENELITIAN

2.1 Analisis Masalah

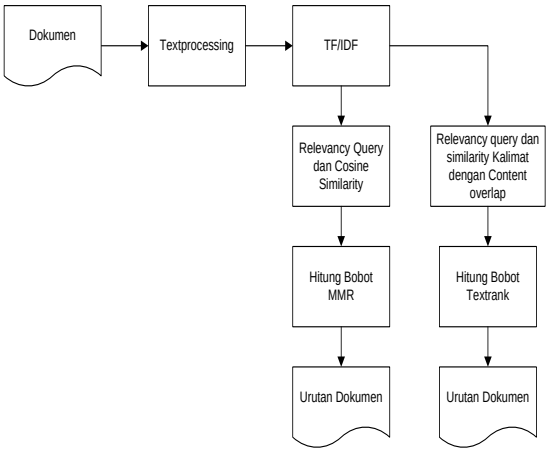
Metode MMR dan metode *textrank* merupakan ektraksi dokumen yang sangat efektif pada pencarian dokumen. Masalah pada pencarian dokumen adalah bagaimana mengetahui keakurasian pada pencarian dokumen yang relevan.

Pada ektraksi dokumen dengan metode MMR dilakukan tokenizing dokumen menjadi kalimat dan dilakukan pengelompokan sesuai dengan gender kalimat tersebut. MMR dilakukan dengan kombinasi matrik cosine similarity untuk merangking kalimat-kalimat sesuai dengan tanggapan yang diberikan oleh user. Pada penelitian yang dilakukan oleh muchammad mustaqhfiri keakurasian yang dihasilkan oleh metode MMR dengan persentase 100% dengan nilai tertinggi dan 50% dengan nilai terendah[1].

Pada ektraksi dokumen yang dilakukan oleh yuzar marsyah pada penelitiannya kekakurasian yang dihasilkan dengan persentase teritngi 95.56%, sehingga perlu diteliti keakurasian pada perbandingan metode *maximum marginal relevance* dan algoritma *textrank* pada pencarian dokumen.

2.2 Analisis Sistem Yang Akan Dibangun

Gambaran dari tahapan yang akan dilakukan sistem yaitu pada Gambar 1.



Gambar 1 Gambaran Umum Sistem

2.3 Analisis Text Processing

Diberikan sebuah dokumen berita yang akan d analisis oleh metode MMR dan textrank. Contoh dokumen sebagai berikut:

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.

Contoh dokumen tersebut akan di proses ketahapan selanjutnya.

1. Tokenizing kalimat, memisahkan kalimat berdasarkan delimeter yang menghubungkannya seperti titik.

Tabel 1 Tokenizing Kalimat

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* merupakan Proses mengubah huruf kapital menjadi huruf kecil dan membuang semua tanda baca, angka dan simbol.

Tabel 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,* membuang kata" yang kurang penting atau kata yang tidak mempunyai makna.
4. 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

5. *Stemming,* mengembalikan suatu kata ke bentuk akarnya (*root word*) dengan aturan-aturan tertentu, sehingga setiap kata memiliki representasi yang sama.

Tabel 4 *Stemming*

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

6. Pembobotan TF/IDF, kemunculan suatu *term* dalam sebuah dokumen *term frequency* (*tf*) dan jumlah kemunculan *term* dalam koleksi dokumen *inverse document frequency* (*idf*).

Tabel 5 Perhitungsn 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

Tabel 6 Perhitungan Bobot 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

Setelah itu, lakukan penjumlahan keseluruhan bobot term pada segmen tertentu, pada segmen S1 sebagai berikut.

$$\sum W(TF) = 14.876$$

Keterangan:

TF(Term Frequency): Jumlah Kemunculan kalimat

Tabel 7 Bobot 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

Selanjutnya perkalian antara query dengan setiap dokumen (segmentasi kalimat), pada segmen S1 berikut.

$$\Sigma(W(Q)*W(S1))= 0.065$$

Tahap kedua adalah dengan menghitung panjang kalimat dalam setiap dokumen termasuk query dengan melakukan pengkuadratan terhadap bobot term setiap dokumen, lakukan penjumlahan kuadrat kemudian diakarkan sehingga menghasil penyebut dalam persamaan.

$\sqrt{\sum(TF)^2}$ pada segmen 1 atau kalimat 1 adalah

$$\sqrt{11.826} = 3.439$$

Maka diperoleh hasil keseluruhan seperti berikut.

Tabel 8 Panjang Kalimat

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

Setelah itu, terapkan rumus cosine similarity untuk menghitung suatu kemiripan query dengan kalimat lainnya. Kita ambil contoh pada S6, sebagai berikut.

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

Maka diperoleh hasil keseluruhan seperti berikut.

Tabel 9 Nilai Hasil Kemiripan Antar Kalimat

	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

Algoritma *maximum marginal relevance* digunakan untuk merangking kalimat-kalimat sebagai tanggapan terhadap *query* yang diberikan oleh user. Perhitungan MMR dilakukan dengan iterasi yaitu *query* terhadap keseluruhan kalimat dan *similarity* antara kalimat dengan kalimat.

Pengguna yang menginginkan informasi ruang sampel disekitar *query*, maka harus menetapkan λ pada nilai yang lebih rendah. Sedangkan bagi pengguna yang ingin fokus untuk memperkuat dokumen-dokumen yang agar lebih relevan, maka harus menetapkan λ pada nilai yang lebih dekat dengan λ[1].

Kalimat dengan nilai MMR tertinggi dari setiap perhitungan iterasi akan di ambil, yang kemudian dipilih untuk dijadikan sebagai ektraksi dokumen. Iterasi akan berhenti ketika hasil mmr maksimum sama dengan 0.

MMR

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

Tabel 10 iterasi 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	-	-	-

Setelah itu urutkan hasil iterasi keseluruhan dari mulai iterasi awal sampai iterasi akhir dengan mengurutkan hasil tertinggi sampai ke yang terendah.

Tabel 11 Hasil Bobot 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

Hasil nilai tertinggi pembobotan MMR ini adalah S2 dengan bobot 0.3017 yang akan dijadikan untuk menampilkan dokumen yang relevan sesuai query yang dipanggil oleh user.

2.4 Algoritma Textrank

Algoritma textrank adalah algoritma ektraksi kalimat yang dilakukan terhadapa dokumen tunggal maupun multi dokumen. Algoritma textrank merupakan pengolahan kalimat untuk mendapatkan dokumen yang relevan sesuai tanggapan yang diberikan oleh user.

Untuk memisahkan isi dokumen menjadi kalimat-kalimat tunggal dengan menggunakan splitter berupa tanda baca seperti titik("."), tanda seru ("!"), tanda tanya ("?"), dan newline. Setelah melakukan pemisahan kalimat dengan menggunakan splitter, selanjutnya menghitung similarity dengan content overlap. Metode pemeringkatan berbasis graf, vertex mempresentasikan kalimat-kalimat yang berada dalam dokumen sedangkan edge mempresentasikan similarity antara dua kalimat yang saling berelasi jadi setiap kalimat tunggal yang diperoleh dari proses pemisahan kalimat.

Pada tahap ini menghitung bobot relevancy terhadap query yang di inputkan oleh *user*. Setelah itu, lakukan penjumlahan keseluruhan bobot term pada segmen tertentu, pada segmen Q sebagai berikut:

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

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

Pada tahapan ini adalah menghitung kesamaan/similarity antar segmen atau antar kalimat dengan content overlap, dari pengambilan kalimat yang sama terhadap setiap segmen kalimat.

Selanjutnya perkalian antara query dengan setiap dokumen (segmentasi kalimat) pada segmen S2 dengan persamaan 2.4 berikut:

$$\Sigma(W(Q)*W(S2))= 0.228$$

Tabel 13 Perkalian dengan 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

Tahap kedua adalah dengan menghitung panjang setiap dokumen termasuk query dengan melakukan penjumlahan terhadap bobot term setiap dokumen, kemudian di log.

Log(SiSj) pada segmen 2 adalah 5.03= 0.702

Maka diperoleh hasil keseluruhan seperti berikut:

Tabel 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

Setelah itu, terapkan rumus similarity content overlap 2.3, untuk menghitung suatu kemiripan query dengan kalimat lainnya. Contoh pada S2, sebagai berikut:

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

Tabel 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

Proses pembobotan bertujuan untuk mendapatkan bobot akhir dari suatu dokumen. Semakin banyak bobot akhir dari suatu dokumen, maka semakin relevan informasi di dalamnya.

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

Setelah itu urutkan hasil iterasi keseluruhan dari mulai S1 sampai S8 dengan mengurutkan hasil tertinggi sampai ke yang terendah.

Tabel 16 Hasil Iterasi 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

Dari tabel diatas dapat di ambil kalimat ke S1 sebagai nilai tertinggi sebagai sample kalimat yang akan dijadikan untuk menampilkan dokumen yang relevan.

Hasil analisis MMR dan Textrank adalah mengumpulkan semua bobot dari setiap dokumen yang sudah dibobot untuk mengetahui dokumen yang relevan. Pengujian yang dilakukan terkait dengan jumlah dokumen sebanyak 31. Query dimasukan dalam percobaan yang dilakukan adalah BPJS menghasilkan 5 dokumen dengan bobot tertinggi dengan query yang dimasukan oleh user. Dapat dilihat selengkapnya pada tabel.

Tabel 17 Hasil Bobot MMR dan 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 Pengujian

Rencana pengujian yang akan dilakukan pada aplikasi perbandingan algoritma *MMR* dan *Textrank* dibagi menjadi 2 bagian pengujian *blackbox* dan pengujian *confusion matrix*. Pengujian *blackbox* berfokus pada perintah-perintah kondisi yang di eksekusi oleh perangkat lunak dan pengujian ini dilakukan oleh pembangun aplikasi. Pengujian *confusion matrix* berfokus pada hasil yang diproses oleh aplikasi untuk mengukur suatu keakurasian hasil aplikasi.

a. Pengujian blackbox

Berdasarkan hasil pengujian *black box* (fungsional) dengan kasus uji di atas, dapat ditarik kesimpulan bahwa pembangunan aplikasi perbandingan Algoritma *MMR* dan *Textrank* tidak terdapat kesalahan. Pada proses secara fungsional aplikasi yang dibangun mengeluarkan hasil yang sesuai dengan harapan.

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

Tabel 18 Hasil Pengujian Blackbox

b. Pengujian MMR

Pengujian MMR yang dilakukan adalah dengan memasukan query terlebih dahulu dengan contoh query "BPJS" terhadap 35 dokumen berita.

Tabel 19 Pengujian 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

Pada tabel pengujian MMR dengan query "bpjs" menghasilkan pencarian yang benar sebanyak 5 dokumen berita, 30 dokumen berita yang tidak benar, dan tidak adanya hasil yang hilang dalam pencarian dokumen.

Tabel 20 Hasil Confusion Matrix

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

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

c. Pengujian *Textrank*

Tabel pengujian *confusion matrix textrank*, menyatakan hasil pencarian yang benar ada 5

dokumen berita dari total 35 dokumen berita, serta hasil yang tidak sesuai terdapat 30 berita.

Tabel 21 Hasil Pengujian 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

Hasil pengujian dengan query "bpjs", dokumen berita sebanyak 35, menghasilkan nilai benar sebanyak 5 dokumen dan 30 dokumen yang salah.

Tabel 22 Hasil Confusin Matrix Textrank

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

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

Berdasarkan hasil pengujian yang sudah dilakukan dapat ditarik suatu kesimpulan bahwa perangkat lunak berfungsi sesuai yang diharapkan dari awal serta perangkat lunak yang dibangun telah cukup memenuhi tujuan awal pembangunan. Algoritma textrank cenderung lebih lambat dalam pemrosesan dokumen dibanding dengan algoritma MMR.

Pada pengujian dengan query "BPJS" menghasilkan pencarian yang sesuai sebanyak 5 dokumen berita dari total berita 35 dan menghasilkan 30 berita yang tidak sesuai. Keakurasian dari kedua algoritma tersebut sangatlah bagus rata-rata hasil pencarian dokumen mempunyai tingkat keakurasian 100%.

3 PENUTUP

Berdasarkan penelitian mengenai perbandingan algoritma *MMR* dan *Textrank*, penguji berhasil mengetahui keakurasian yang dihasilkan oleh kedua algoritma tersebut. Hasil dari pengujian membuktikan nilai keakurasian yang dihasilkan oleh kedua algoritma menghasilkan keakurasian 100% dengan jumlah dokumen berita sebanyak 35, dan menghasilkan keakurasian sesuai *query* sebanyak 5 dokumen berita dan 30 berita yang tidak sesuai, waktu pencarian yang dihasilkan oleh *MMR* lebih cepat dibandingkan dengan *textrank* sesuai dengan jumlah dokumen yang diproses.

Berdasarkan hasil pengujian yang telah dilakukan, maka saran yang diharapkan setelah dilakukannya analisis dan perbandingan algoritma *MMR* dan algoritma *Textrank.*

1. Pengembangan lebih lanjut memakai *generator* judul, untuk memastikan isi dokumen dan judul sama memiliki keterkaitan, karena pemrosesan dokumen lebih kepada isi kalimat.
2. Disarankan untuk menguji dokumen yang lebih banyak lagi.

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.