

INFORMATION RETRIEVAL (IR) PENCARIAN IDE POKOK DALAM TEKS ARTIKEL OLAHRAGA BERBAHASA INGGRIS MENGUNAKAN METODE MMR (*MAXIMUM MARGINAL RELEVANCE*)

Kamil Malik¹⁾, Moh.Jasri²⁾, Ahmad Sanusi Mashuri³⁾
Prodi Informatika Fakultas Teknik Universitas Nurul Jadid
Karanganyar Paiton Probolinggo

e-mail: nomor1001@gmail.com¹⁾, jasriahyak@gmail.com²⁾, ahmadsanusi260197@gmail.com³⁾

ABSTRAK

Dalam dunia Internet terdapat banyak sekali informasi-informasi berupa artikel yang tersebar luas dalam berbagai bahasa. Salah satu jenis artikel yang banyak dijumpai di dunia maya adalah artikel bahasa Inggris. Artikel berbahasa Inggris merupakan artikel dengan bahasa yang paling cepat menyebar dan terbaru dikarenakan bahasa Inggris merupakan bahasa internasional yang banyak digunakan orang di dunia. Salah satu jenis artikel yang banyak diminati khususnya dikalangan remaja yaitu artikel olahraga. Namun dalam hal ini tidak semua artikel yang dibutuhkan menggunakan bahasa Indonesia, terdapat beberapa artikel yang menggunakan bahasa Inggris. Dengan ini kalangan remaja kesulitan untuk mengetahui intisari atau ide pokok yang terdapat dalam artikel tersebut. Salah satu cara untuk peringkasan yang sudah diterapkan adalah Peringkasan Teks Otomatis (Automated Text Summarization), Tujuannya untuk mengambil sumber informasi dengan mengutip sebagian besar isi yang penting dan menampilkan kepada pembaca dalam bentuk ringkas sesuai dengan kebutuhan pembaca. Dengan demikian teknologi ini dapat membantu pembaca untuk menyerap informasi yang terdapat dalam teks melalui ringkasan tanpa membaca keseluruhan dokumen, metode yang digunakan dalam penelitian ini yaitu MMR (maximum marginal relevance). Berdasarkan dari permasalahan yang timbul dalam penelitian ini, telah dihasilkan peringkasan teks artikel olahraga berbahasa Inggris menggunakan metode MMR (*maximum Marginal Relevance*). Dan yang menjadi ringkasan adalah kalimat 1 (D1) dengan nilai maksimal MMR adalah 0,0962094 dan kalimat 2 (D2) dengan nilai maksimal MMR adalah 0,0928395, dengan nilai akurasi yaitu 75%.

Kata Kunci : MMR, Peringkasan Teks

1. PENDAHULUAN

Dalam dunia Internet terdapat banyak sekali informasi-informasi berupa artikel yang tersebar luas dalam berbagai bahasa. Salah satu jenis artikel yang banyak dijumpai di dunia maya adalah artikel bahasa Inggris. Namun, tidak semua orang fasih berbahasa Inggris dan membutuhkan bantuan untuk mengerti isi, sementara artikel yang akan dibaca sangat penting dan panjang sehingga dapat menyita waktu.

Salah satu jenis artikel yang banyak diminati khususnya dikalangan remaja yaitu artikel olahraga. Namun dalam hal ini tidak semua artikel yang dibutuhkan menggunakan bahasa Indonesia, terdapat beberapa artikel yang menggunakan bahasa Inggris.

Salah satu cara untuk peringkasan yang sudah diterapkan adalah Peringkasan Teks

Otomatis (*Automated Text Summarization*), yaitu sebuah proses untuk menghasilkan ringkasan dari teks menggunakan komputer. Tujuannya untuk mengambil sumber informasi dengan mengutip sebagian besar isi yang penting dan menampilkan kepada pembaca dalam bentuk ringkas sesuai dengan kebutuhan pembaca.

Basis pertama yang digunakan adalah *Term Frequency Inverse Document Frequency* (TF-IDF). Beberapa metode yang sering digunakan diantaranya adalah metode *genetika*, *lexrank*, *maximum marginal relevance*, *graph based*. Namun yang digunakan dalam penelitian ini adalah metode maximum marginal relevance

2. STUDI PUSTAKA

2.1 Information retrieval

Sistem temu kembali informasi (IR) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari penggunaanya secara otomatis berdasarkan kesesuaian dengan query dari suatu koleksi informasi (Karmasya & Mahendra, 2012). Prinsip kerja sistem temu kembali informasi jika ada sebuah kumpulan dokumen dan seorang user yang memformulasikan sebuah pertanyaan (request atau query). Jawaban dari pertanyaan tersebut adalah sekumpulan dokumen yang relevan dan membuang dokumen yang tidak relevan (Salton, 1989).

2.2 TF-IDF (*Term Frequency-Inverse Document Frequency*)

Basis pembobotan TF-IDF adalah jenis pembobotan yang sering digunakan dalam IR dan *text mining*. Pembobotan ini adalah suatu pengukuran statistik untuk mengukur seberapa penting sebuah kata dalam kumpulan dokumen. Tingkat kepentingan meningkat ketika sebuah kata muncul beberapa kali dalam sebuah dokumen tetapi diimbangi dengan frekuensi kemunculan kata tersebut dalam kumpulan dokumen. TF-IDF dapat dirumuskan sebagai berikut,

$$TF - IDF (t_{k,d_j}) = (t_{k,d_j}) * IDF (t_k) \dots \dots \dots (2,1)$$

Keterangan :

d_j = dokumen ke-j

t_k = term ke-k

Dimana sebelumnya dihitung terlebih dahulu *Term Frequency* (TF) yaitu frekuensi kemunculan suatu *term* di tiap dokumen. Kemudian dihitung *Inverse Document Frequency* (IDF) yaitu nilai bobot suatu term dihitung dari seringnya suatu term muncul di beberapa dokumen. Semakin sering suatu term muncul di banyak dokumen, maka nilai IDF nya akan kecil. Berikut rumus-rumus TF dan IDF.

$$TF (t_{k,d_j}) = F (t_{k,d_j}) \dots \dots \dots (2,2)$$

Keterangan :

TF = Jumlah frekuensi term

f = Jumlah frekuensi kemunculan f

d_j = dokumen ke-j

t_k = term ke-k

Lalu untuk menghitung nilai IDF bisa menggunakan persamaan sebagai berikut, yaitu :

$$IDF (t_k) = \frac{1}{df (t)} \dots \dots \dots (2,3)$$

Atau

$$IDF (t_k) = \log \frac{1}{df (t)} \dots \dots \dots (2,4)$$

Keterangan :

IDF = bobot term

N = jumlah total dokumen

d_j = dokumen ke-j

t_k = term ke-k

2.3 Cosine Similarity

Algoritma ini digunakan untuk menghitung kesamaan antar 2 buah objek yang dinyatakan dalam 2 buah vektor dengan menggunakan keyword dari sebuah dokumen sebagai ukuran.

2.4 MMR

Maximum Marginal Relevance (MMR) merupakan salah satu metode ekstraksi ringkasan (*extractive summary*) yang digunakan untuk meringkas dokumen tunggal atau multi dokumen. MMR meringkas dokumen dengan menghitung kesamaan antara bagian teks dan dengan tujuan mendapatkan skor kalimat berdasarkan kesamaan (*similarity*) dengan *query* yang diberikan dan dapat mengurangi redudansi pada hasil ringkasan yang di dapat (Goldstein, 2008).

3. METODE PENELITIAN

Metodologi penelitian digunakan sebagai kerangka kerja dalam penelitian karya ilmiah sesuai dengan hasil yang diharapkan dan tidak menyimpang dari tujuan yang dirumuskan sebelumnya. Berikut merupakan penjelasan dari metodologi penelitian :



Gambar 3.1 Tahapan Penelitian

3.1. Prosedur Penelitian

Berdasarkan model pengembangan yang digunakan secara garis besar model waterfall mempunyai langkah-langkah sebagai berikut :

a. Perumusan Masalah

Merumuskan masalah tentang topik atau tema penelitian yang akan dibahas pada tugas akhir ini.

b. Pengumpulan Data

Tahap pengumpulan data dilakukan untuk mendapatkan informasi informasi dan data-data yang terkait dalam penelitian tugas akhir ini. Adapun aktivitas yang telah dilakukan pada tahapan pengumpulan data, yaitu:

1. Studi Pustaka

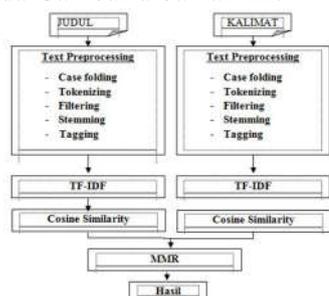
Untuk mendapatkan dasar-dasar pengetahuan yang akan diterapkan dalam penelitian ini dan memperoleh informasi pada penelitian ini, maka dipelajari bahan pustaka dan referensi-referensi dari sumber sumber terkait seperti buku cetak, website, ebook, jurnal, artikel artikel, dan sumber-sumber informasi lainnya yang berhubungan dengan metode penelitian yang meliputi konsep metode term freuency-inverse document frequency, cosine similarity dan MMR.

2. Diskusi

Pendekatan lain yang dilakukan pada tahap ini adalah diskusi dan mempelajari masalah-masalah pada penelitian tugas akhir. Penulis berdiskusi dengan dosen pembimbing dan peneliti yang telah melakukan penelitian yang membahas tentang peringkasan Artikel

c. Analisa

Terdapat beberapa tahapan yang dilakukan untuk mempermudah pencarian ide pokok dalam teks artikel olahraga berbahasa inggris. Adapun alur dari tahapan-tahapan tersebut ditunjukkan pada Gambar dibawah ini.



Gambar tahapan text processing

d. Perancangan

Setelah melakukan analisa, maka kemudian dilanjutkan dengan perancangan sistem berdasarkan analisa permasalahan yang telah dilakukan sebelumnya. Pada perancangan perangkat lunak ini mencakup:

1. Perancangan subsistem model adalah merancang flowchart sistem
2. Perancangan subsistem dialog adalah merancang tampilan antar muka sistem (user interface) dan struktur menu.
3. Perancangan procedural merupakan tahap perancangan pada metode atau algoritma yang akan digunakan dalam membangun sistem.

e. Implementasi

Implementasi dan pengujian adalah metode terakhir yang dilakukan setelah tahap analisa dan perancangan selesai. Tahap implementasi ini menjelaskan tentang penerapan pengujian berhasil atau tidaknya penelitian yang dilakukan.

f. Pengujian

Pengujian dilakukan dengan menggunakan 4 artikel olahraga berbahasa inggris yang diambil dari beberapa media internet (bbc.co.uk/sport,global.espn.com, marca.com/en).data yang akan diuji adalah sebagai berikut, yaitu:

Ringkasan manual Artikel Olahraga berbahasa inggris

No	Judul	Ide Pokok
1.	Belgium's curious success across continents at the World Cup	belgium were one of the standout teams at this year's World Cup in Russia and have left after achieving a new record. For the first time in the competition's history, the Red Devils can boast of having won a game against opponents from five different confederations in a single edition of the World Cup. The five confederations from which Belgium have beaten a team : Panama (CONCACAF) ,Tunisia (CAF), England (UEFA), Japan (AFC) ,Brazil (CONMEBOL). Therefore, they leave Russia with several records to their credit since, in addition to being a side from each confederation, they have also achieved their best ever finish with a bronze medal

4. HASIL DAN PEMBAHASAN

4.1. Analisis Data

Pada bagian ini akan dijelaskan hasil pengujian dan analisis pengaruh metode

MMR dan TF-IDF-DF terhadap hasil ringkasan. Pengujian dilakukan dengan membandingkan hasil ringkasan oleh sistem dengan ringkasan referensi (ringkasan manual).

1. Hasil Pengujian Manual

Hasil peringkasan teks manual yang dilakukan oleh peneliti tentang artikel olahraga berbahasa Inggris yang bersumberkan dari (bbc.co.uk/sport, global.espn.com, marca.com/en). Hasil ringkasan manual ditunjukkan pada tabel dibawah ini

No	Judul	Ide Pokok
1.	Belgium's curious success across continents at the World Cup	belgium were one of the standout teams at this year's World Cup in Russia and have left after achieving a new record. For the first time in the competition's history, the Red Devils can boast of having won a game against opponents from five different confederations in a single edition of the World Cup. The five confederations from which Belgium have beaten a team : Panama (CONCACAF) ,Tunisia (CAF), England (UEFA), Japan (AFC) ,Brazil (CONMEBOL). Therefore, they leave Russia with several records to their credit since, in addition to being a side from each confederation, they have also achieved their best ever finish with a bronze medal

2. Hasil Pengujian sistem

Peringkasan teks dengan sistem dilakukan dengan cara menginput isi artikel ke dalam sistem. Selanjutnya, sistem memproses inputan isi artikel, sehingga sistem akan mengeluarkan hasil ringkasan. Salah satu hasil peringkasan teks otomatis yang dilakukan oleh sistem terlihat pada tabel 4.2

No	Judul	Ide Pokok
1.	Belgium's curious success across continents at the World Cup	they leave Russia with several records to their credit since, in addition to being a side from each confederation, they have also achieved their best ever finish with a bronze medal run and been the tournament's top scorers with 16. Even in the final he stepped up and netted for a side which are fondly remembered as one of the best footballing teams of all time, led by the legendary Pele. If there is a record that nobody has been able to beat or even equal in the tournament's history then it is the one achieved by Jair Ventura Filho, popularly known as Jairzinho, in the 1970 World Cup in Mexico.

Untuk memperoleh hasil ringkasan artikel olahraga seperti tabel di atas sistem harus melalui beberapa tahap seperti *text preprocessing*, pembobotan

kata dengan metode TF-IDF-DF, menghitung *cosine similarity* dan MMR.

a. Text Processing

Tujuan dari tahap *text preprocessing* yaitu merubah artikel olahraga menjadi kata-kata yang siap diproses untuk perhitungan bobot kata. Beberapa proses dari *text preprocessing*, yaitu segmentasi kalimat, *case folding*, *tokenizing*, *filtering*, *stemming* dan *tagging*. Berikut ini adalah salah satu contoh dokumen yang diinputkan dalam proses *text preprocessing* disertai tahapan proses *text preprocessing* :

1.) Segmentasi Kalimat

Merupakan pemecahan paragraf menjadi kalimat. Pemecahan dilakukan dengan memisahkan berdasarkan tanda baca titik (.), tanda tanya (?) dan tanda seru (!).

2.) Case folding

Paragraf artikel olahraga yang telah dipotong menjadi kalimat akan menjalankan proses *case folding*. *Case folding* adalah mengubah semua teks menjadi karakter dengan huruf kecil dan membuang semua karakter selain a-z. Selain itu, tanda baca, bilangan numerik dan simbol juga dihilangkan.

3.) Tokenizing

Merupakan proses pemotongan kalimat menjadi kata-kata. Pemotongan kalimat berdasarkan *delimiter* yang menyusunnya, yaitu spasi (" ").

4.) Filtering

Dalam tahap filtering ini melakukan pembuangan stopword. Stopword adalah kata-kata yang tidak memiliki makna atau kata yang kurang berarti dan sering muncul dalam kumpulan kata. Cara membuang kata-kata yang tidak penting dengan mengecek pada kamus stopword. Jika kata ada yang sama dengan stopword, maka kata akan dibuang atau dihapus.

5.) Stemming

Mengembalikan kata-kata yang diperoleh dari hasil filtering kebentuk dasarnya, menghilangkan imbuhan awal (*prefix*) dan imbuhan akhir (*suffix*) sehingga di dapat kata dasar. Metode stemming memerlukan masukan berupa kata yang terdapat dalam suatu

dokumen, dengan menghasilkan keluaran berupa root word.

6.) *Tagging*

Merubah kata dalam bentuk lampau (past tense) menjadi bentuk sekarang (present tense).

b. Algoritma TF-IDF

Setelah proses *text preprocessing*, tahap selanjutnya yaitu penghitungan bobot kata dengan algoritma TF-IDF-DF. Berikut adalah proses perhitungan bobot kata.

$$w_{i,j} = \left(tf_{i,j} \times \log \left(\frac{N}{df_i} \right) \right) \times df_i$$

Untuk hasil perhitungan bobot kata pada semua *term* dengan metode TF-IDF-DF, dapat dilihat pada gambar dibawah ini:

Term	Tf				D	idf	W				
	Q	D1	D2	D3			D4	Q	D1	D2	D3
Messi	1	6	7	1	0	3	0.7826	4.6956	5.4782	0.7826	0
Head	0	1	1	0	0	2	1.2041	1.2041	0	0	0
Ballon	0	3	0	0	0	1	2.246	7.2246	0	0	0
D'or	0	3	0	0	0	1	2.4082	7.2246	0	0	0
Leo	1	2	1	1	0	3	0.7826	1.5652	0.7826	0.7826	0

Keterangan:

Q : *query*

tf : *term frequency*

df : *document frequency*

idf : *inverse document frequency* (Log10(N/DF))

W : Bobot Kata (TF * IDF * DF)

c. *Cosine similarity*

Jika bobot kata telah diperoleh, selanjutnya mencari nilai *cosine similarity*. Perhitungan *cosine similarity* dibagi menjadi 2 tahap, yaitu :

1. Perhitungan relevansi antara dokumen dan *query* (judul) Menghitung cosinus sudut dari dua vektor, yaitu W (bobot) dari tiap dokumen atau kalimat dengan W (bobot) dari *query* (judul).
2. Perhitungan *similarity* antara dokumen Menghitung *cosine* sudut vektor W (bobot) suatu kalimat dengan vektor W (bobot)

kalimat yang lain.. Hasil perhitungan *cosine similarity* dapat dilihat pada gambar dibawah ini:

Wds * Wdi					panjang vector				
D1	D2	D3	D4	Q	D1	D2	D3	D4	
3.674777	4.287239	0.612463	0	0.612463	22.04866	30.01068	0.612463	0	0
0	0	0	0	0	1.449857	1.449857	0	0	0
0	0	0	0	0	52.19485	0	0	0	0
0	0	0	0	0	52.19485	0	0	0	0
1.224926	0.612463	0.612463	0	0.612463	2.449851	0.612463	0.612463	0	0

Pada gambar di atas adalah perhitungan *Cosine Similarity*, Langkah 2c: terapkan rumus cosine similarity. Hitung kemiripan Q dengan D1, D2, D3, D4.

Cos (Q, D1) =

$$6.349318 / (1.742753 * 26.50767) = 0.137442$$

Cos (Q, D2) =

$$6.711963 / (1.742753 * 29.02854) = 0.132675$$

Cos (Q, D3) =

$$3.39959 / (1.742753 * 22.82649) = 0.085458$$

Cos (Q, D4) =

$$0.362404 / (1.742753 * 29.80124) = 0.006978$$

Gambar relevansi antara judul dan kalimat

D1	D2	D3	D4
0.137442	0.132675	0.085458	0.006978

Gambar similarity antar kalimat

	D1	D2	D3	D4
D1	0.00142	0.00011	5.38706	7.29656
D2	0.00011	0.011867	4.51774	3.88979
D3	5.38706	4.51774	0.00191	4.31344
D4	7.29656	3.88979	4.31344	0.00112

d. MMR

Setelah perhitungan *cosine similarity* diperoleh, maka tahap berikutnya menghitung nilai MMR. Algoritma *maximum marginal relevance* digunakan untuk meranking kalimat-kalimat sebagai tanggapan terhadap *query* yang diberikan user. Perhitungan MMR dilakukan dengan iterasi mengkombinasikan 2 matrik *cosine similarity*, yaitu relevansi antara *query* terhadap keseluruhan kalimat dan *similarity* antara kalimat.

1. Perhitungan iterasi ke-1

$$\begin{aligned} \text{MMR} (1) &= 0,7 * \text{Sim}_1 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_1, S') \\ &= 0,7 * 0.137442 - (1 - 0,7) * 0 = \\ &= 0,0962094 \end{aligned}$$

$$\begin{aligned} \text{MMR} (2) &= 0,7 * \text{Sim}_2 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_2, S') \\ &= 0,7 * 0.132675 - (1 - 0,7) * 0 = \\ &= 0,0928725 \end{aligned}$$

$$\begin{aligned} \text{MMR} (3) &= 0,7 * \text{Sim}_3 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_3, S') \\ &= 0,7 * 0.085458 - (1 - 0,7) * 0 = \\ &= 0,0598206 \end{aligned}$$

$$\begin{aligned} \text{MMR} (4) &= 0,7 * \text{Sim}_4 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_4, S') \\ &= 0,7 * 0.006978 - (1 - 0,7) * 0 = \\ &= 0,0048846 \end{aligned}$$

Dari hasil perhitungan pada iterasi ke-1, diperoleh nilai maximum MMR = 0,0962094

2. Perhitungan iterasi ke-2

$$\begin{aligned} \text{MMR} (2) &= 0,7 * \text{Sim}_2 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_2, S_1) \\ &= 0,7 * 0.132675 - (1 - 0,7) * 0.00011 \\ &= 0,0928395 \end{aligned}$$

$$\begin{aligned} \text{MMR} (3) &= 0,7 * \text{Sim}_3 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_3, S_1) \\ &= 0,7 * 0.085458 - (1 - 0,7) * 5.38706 \\ &= -1,5562974 \end{aligned}$$

$$\begin{aligned} \text{MMR} (4) &= 0,7 * \text{Sim}_4 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_4, S_1) \\ &= 0,7 * 0.006978 - (1 - 0,7) * 7.29656 \\ &= -2,1840834 \end{aligned}$$

Dari hasil perhitungan pada iterasi ke-2, diperoleh nilai maximum MMR = 0,0928395

3. Perhitungan iterasi ke-3

$$\begin{aligned} \text{MMR} (2) &= 0,7 * \text{Sim}_2 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_2, S_1) \\ &= 0,7 * 0.132675 - (1 - 0,7) * 0.00011 \\ &= 0,0928395 \end{aligned}$$

$$\begin{aligned} \text{MMR} (3) &= 0,7 * \text{Sim}_3 (S_i, Q) - (1 - 0,7) * \text{maxsim}_2(S_3, S_1) \\ &= 0,7 * 0.085458 - (1 - 0,7) * 5.38706 \\ &= -1,5562974 \end{aligned}$$

Dari hasil perhitungan pada iterasi ke-3, diperoleh nilai maximum MMR < 0 pada D2 dan D3, sehingga tidak ada kalimat yang dipilih sebagai ringkasan dan iterasi perhitungan akan berhenti Hasil perhitungan iterasi.

	D1	D2	D3	D4
Iterasi ke-1	0,0962094	0,0928725	0,0598206	0,0048846
Iterasi ke-2	0,0928395	-1,5562974	-2,1840834	
Iterasi ke-3	0,0928395	-1,5562974		
Iterasi ke-4				

Pada gambar di atas dipaparkan bahwa pada iterasi 1 kalimat (dokumen) yang tertinggi terdapat pada kalimat 1 (D1), sehingga kalimat 1 (D1) menjadi ringkasan. Pada iterasi 2, yang menjadi ringkasannya berikutnya adalah kalimat 2 (D2). Kalimat 1 (D1) pada iterasi 2 tidak terdapat nilai MMR, karena kalimat 1 (D1) telah dipilih sebelumnya menjadi ringkasan. Iterasi 3 tidak ada kalimat yang menjadi ringkasan, karena nilai max MMR <= 0

Peringkatan hasil ringkasan

Rangking	Kalimat (D) ke	Max MMR
1	1	0,0962094
2	2	0,0928395

5. KESIMPULAN DAN SARAN

5.1 kesimpulan

Berdasarkan dari permasalahan yang timbul dalam penelitian ini, telah dihasilkan peringkasan teks artikel olahraga berbahasa inggris menggunakan metode MMR (*maximum Marginal Relevance*). Dan yang menjadi ringkasan adalah kalimat 1 (D1) dengan nilai maksimal MMR adalah 0,0962094 dan kalimat 2 (D2) dengan nilai maksimal MMR adalah 0,0928395, dengan nilai akurasi yaitu 75%. Menilai dari tabel pengujian bahwa metode MMR layak untuk digunakan dalam *text summarization*.

5.2 Saran

Setelah penelitian ini selesai dan diimplementasikan untuk pengguna umum ada beberapa saran untuk dijadikan acuan terhadap penelitian ini ini ialah sebagai berikut.

1. Disarankan untuk penelitian selanjutnya menggunakan bahasa pemrograman WEB/ Android sebagai pengujian sistem
2. penelitian tentang peringkasan teks ini masih dibutuhkan pengembangan agar penelitian ini dapat menjadi lebih baik dari sekarang. Beberapa kekurangan dari penelitian ini yang dapat dikembangkan adalah :
 - a. Aplikasi yang digunakan untuk pengujian hanya menggunakan satu form dan tidak terkoneksi ke database, bagaimana untuk

penelitian selanjutnya agar menggunakan banyak form sehingga tampilannya lebih bagus dan menarik dan tersimpan ke database.

- b. Aplikasi ini hanya bisa meringkas artikel berbahasa inggris, disarankan untuk penelitian selanjutnya menggunakan banyak bahasa seperti, bahasa arab, dll.
- c. Hanya menggunakan satu metode dalam penelitian ini, bagaimana untuk penelitian selanjutnya menggunakan banyak metode

DAFTAR PUSTAKA

- Yang Guangbing, dkk. (2013). Enhancing Sentence Ordering by Hierarchical Topic Modeling for Multi-document Summarization.
- Baralis Elena, dkk.(2013). Multi-document summarization based on the Yago ontology.
- AL-Khassawneh Yazan Alaya, dkk (2016). Sentence Similarity Techniques for Automatic Text Summarization.
- Achmad ridok, (2014). Peringkasan dokumen bahasa indonesia berbasis Non-Negative Matrix Factorization.
- Winata Fernando, Rainarli Ednawati, (2016). Implementasi Cross Method Latent Semantic analysis untuk meringkas dokumen berita bahasa indonesia.
- Kesuma Hirroe Wijaya Ani, Pribadi Feddy Setio, (2016). Penerapan cosine similarity dalam aplikasi kitab undang-undang hukum dagang (*wetboek van koophandle voor indonesia*).
- Saraswati Nirmala Fa'izah, Indriati, Perdana Rizal Setya, (2018). Peringkasan teks otomatis menggunakan metode *maximum marginal relevance* pada hasil pencarian sistem temu kembali informasi untuk artikel berbahasa indonesia.
- Najibullah Ahmad, Mingyan Wan, (2015). Otomatisasi peringkasan dokumen sebagai pendukung sistem manajemen surat.
- Amin Fatkhul, Anwar Sariyun Naja, Purwaningtyas, (2017). Sistem temu kembali informasi (Information Retrieval System) dokumen berbahasa jawa menggunakan metode *Kumar-Hassebrook (PCE)*
- Anung, G , & Purnama, B. E. (2013).Rancang bangun sistem pengolahan data Gurudan siswa pada madrasah Ibtidaiyah Muhammadiyah (MIM) Glinggangan Anung Gestriawan, Bambang Eka Purnama, 5(3), 7-14.
- Beny, H. (2012). Definisi Information Retrieval. Retrieved from informationretrievalsystem.blogspot.com/2012/08/definisi-information-retrieval.html?m=1
- Staf. (2019). Kata Pencarian Menurut KBBI. Retrieved from <https://www.apaarti.com/pencarian.html>
- Adi Dwi, H. (2017). Menentukan Ide Pokok Paragraf. Retrieved from <https://dapurimajinasi.blogspot.com/2017/09/menentukan-ide-pokok-paragraf.html?m=1>
- Inggris, sekolah bahasa. (2014). Apa Sih Pengertian Dari Bahasa Inggris Itu? Retrieved from <https://www.sekolahbahasainggris.co.id/apa-sih-pengertian-dari-bahasa-inggris-itu/>