

Peringkasan Otomatis Makalah Menggunakan Maximum Marginal Relevance

Wildan Pratama¹, Ridwan Ilyas², Fatan Kasyidi³

^{1,2,3}Jurusan Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia

e-mail: wildan.pratama@student.unjani.ac.id¹, rdwnilyas@gmail.com², fatan.kasyidi@lecture.unjani.ac.id³

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi : 25 Maret 2021

Revisi Akhir : 28 November 2021

Diterbitkan Online : 03 Januari 2022

Kata Kunci :

Peringkasan, Maksimum Marginal Relevance, Membaca, Makalah

Korespondensi :

Telepon / Hp : +62 812-9557-4385

E-mail

wildan.pratama@student.unjani.ac.id

A B S T R A K

Makalah merupakan karya tulis yang memuat pemikiran tentang suatu masalah atau topik tertentu yang ditulis secara sistematis disertai dengan analisis yang logis objektif. Membaca merupakan proses melihat serta memahami isi dari apa yang tertulis. Proses membaca dipengaruhi beberapa factor yang diantaranya jumlah dari bacaan, sehingga penyederhaan atau peringkasan suatu bacaan contoh seperti makalah maka akan menambah kecepatan membaca. Ringkasan merupakan suatu cara yang efektif untuk menyajikan suatu karangan yang panjang dalam bentuk yang singkat. Peringkasan makalah dilakukan untuk menyajikan jumlah bacaan lebih sedikit tanpa mengurangi informasi yang disampaikan dan mempertahankan dalam bentuknya yang singkat. Dalam membuat ringkasan kita diharuskan untuk membaca keseluruhan isi makalah tersebut terlebih dahulu, untuk kemudian meringkasnya. Hal ini tentu menjadi masalah dimana ringkasan dibuat dengan tujuan untuk meminimalkan waktu membaca tetapi tetap dapat memberikan teks yang isinya langsung mengarah pada tujuan utama atau ide pokoknya. Untuk memecahkan masalah tersebut diperlukan suatu perangkat atau aplikasi yang dapat meringkas teks secara otomatis. Beberapa penelitian terdahulu menggunakan graph-based summarization untuk meringkas dokumen tunggal bahasa Indonesia. Penelitian ini membangun sistem yang dapat meringkaskan makalah menjadi ringkasan menggunakan maximum marginal relevance. maximum marginal relevance dipilih karena mampu menilai kalimat dan merankingnya sebagai acuan untuk pembentukan ringkasan. Penelitian ini mendapatkan hasil ringkasan terbaik dengan pengujian Rouge-1 dengan rata-rata yang diraih 0.68.

1. PENDAHULUAN

Ringkasan merupakan penyajian kembali dari suatu tulisan dengan hal-hal penting yang terkandung daritulis tersebut. Pembuatan makalah ilmiah memerlukan informasi dari penelian sebelumnya. Sehingga perlu ada proses pengambilan informasi, yang dimana harus membaca seluruh isi makalah. Proses tersebut akan terbatu jika ada sebuah ringkasan dari makalah tersebut baik ringkasan secara keseluruhan atau perbagian. Natural Language Processing (NLP) adalah area penelitian dan aplikasi yang mengeksplorasi caranya komputer dapat digunakan untuk memahami dan memanipulasi teks atau ucapan bahasa alami agar bermanfaat. NLP bertujuan untuk mengumpulkan pengetahuan tentang bagaimana manusia memahami dan menggunakan bahasa sehingga alat dan teknik yang tepat dapat dikembangkan untuk membuat sistem computer memahami dan memanipulasi bahasa alami untuk melakukan tugas yang diinginkan. Peringkasan teks otomatis dapat diklasifikasikan menjadi dua kategori: ekstraksi dan abstraksi [1]. Perangkuman teks dengan menyalin informasi yang ada dianggap yang paling penting dari teks asli sebagai ringkasan, seperti klausa utama, kalimat utama, atau utama gugus kalimat. Teknik abstraksi mengubah kalimat menjadi kalimat pendek baru yang tidak ada dalam teks asli [2]. Natural Language Processing (NLP) banyak digunakan dalam penelitian yang berkaitan dengan bahasa baik secara tulisan maupun tulisan. Beberapa penelitian tentang NLP, Peneltian sebelumnya melakukan pengelompokan,

ekstraksi informasi dan peringkasan dalam tahap penelitiannya [3]. Penelitian sebelumnya menggunakan metode CWPC (Combined Words Recognition and Paragraph Clustering) menghasilkan recall 77.7% , precision 74.3% dan menghasilkan fmeasure sebesar 72.5% untuk peringkasan berita [4], penelitian lain tentang perigkasan teks otomatis menggunakan metode MMR dengan hasil waktu recall 69%, precision 69% dan menghasilkan fmeasure sebesar 69%, dengan 10 kali 2 pengujian dan banyak data yai 10 berita [5], penelitian sebelumnya menggunakan peringkasan ekstraktif dengan library python NLTK menghasilkan hasil ringkasan lebih baikpada dokumen 1 dengan hasil Rouge-1 F1 0.67 dari hasil ringkasan microsoft word dengan hasil Rouge-1 F1 0.57 [6]. Penelitian ini membuat sistem yang dapat membuat ringkasan dari makalah ilmiah secara otomatis dengan masukan sebuah makalah, akan terangkum menjadi bagian kecil dari teks secara keseluruhan dan mengandung inti, atau hal-hal penting.

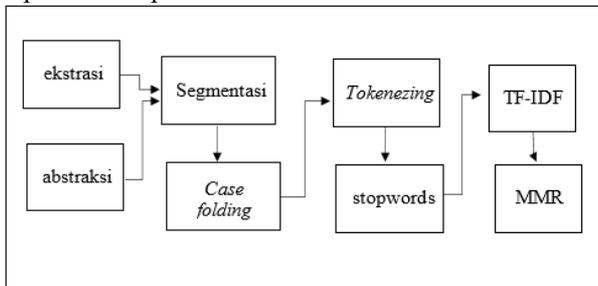
Berdasarkan pada uraian yang dijelaskan sebelumnya dapat diketahui bahwa dokumen makalah terdiri atas struktur data dalam bentuk string yang terdiri kalimat dan kata. Terdapat banyak kemungkinan pemilihan kalimat ringkasan yang ada dalam makalah sehingga menampilkan informasi yang utama. Sehingga, diperlukan algoritma untuk mengolah data dalam bentuk teks dalam satuan kata dan kalimat dan diperlukan algoritma yang mampu menseleksi kalimat-kalimat penting dalam makalah untuk membuat ringkasan. Oleh

karena itu dalam penelitian ini akan dibuat perangkat lunak untuk memproses data teks dan algoritma MMR untuk memilih kalimat-kalimat yang mengandung informasi utama.

Berdasarkan pada permasalahan yang telah diuraikan, tujuan dari penelitian adalah membangun sistem peringkasan teks otomatis untuk makalah ilmiah menggunakan TF-IDF dan MMR. Tujuan lainnya adalah untuk memahami penggunaan TF-IDF dan MMR dalam peringkasan teks pada makalah ilmiah.

2. METODE PENELITIAN

Tahapan dari metodologi penelitian untuk peringkasan teks otomatis pada makalah ilmiah menggunakan Term Frequency - Inverse Document Frequency (TF-IDF) dengan MMR. Tahapan penelitian ini terdiri dari tiga tahapan yaitu proses pra proses, proses, dan hasil. Tahapan metodologi penelitian diperlihatkan pada Gambar 1.



Gambar 1. Tahapan Metodologi Penelitian

Pada Gambar 1.1 menjelaskan skema dari proses peringkasan teks secara otomatis. Tahap awal pada peringkasan teks adalah melakukan ekstraksi, menyalin unit-unit teks pada makalah. Tahap abstraksi terhadap makalah. Pembentukan kalimat hasil dari proses abstraksi, melakukan pemerataan case. Selanjutnya pemisahan perkata, menghapus kata tidak perlu.

3. TINJAUAN PUSTAKA

Terdapat beberapa pengertian yang masuk kedalam daftar tinjauan pustaka sebagai berikut :

3.1. Peringkasan Teks Otomatis

Peringkasan teks otomatis adalah proses menghasilkan atau mengekstraksi teks dari satu atau lebih teks yang berisi informasi penting dari teks asli, dan teks yang dihasilkan secara otomatis tidak melebihi setengah teks asli.. Peringkasan dibagi menjadi dua jenis: pertama peringkasan ekstraktif, kedua peringkasan abstraktif. Peringkasan teks Ekstraktif memilih kalimat yang diperlukan dari teks. Paragraf penting ini dapat dipilih dengan menggunakan fitur linguistik dan statistik paragraf. Peringkasan abstraktif mengerti yang utama konsep dan makna dokumen atau teks email yang diberikan [7].

3.2. Pendekatan Peringkasan Teks

Pendekatan peringkasan teks ada dua [8], pertama peringkasan ekstraksi, dan peringkasan abstraksi :

A. Ekstraksi

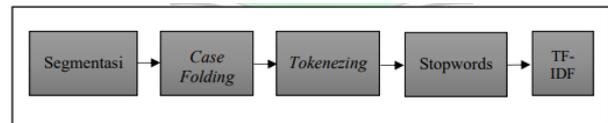
Ekstraksi abstrak adalah proses menyalin blok teks terpenting dari teks Anda ke dalam resume Anda. Blok teks yang disalin dapat berupa klausa utama, klausa utama, atau klausa utama tanpa menambahkan 10 kalimat baru yang ditemukan dalam dokumen asli. Pada penelitian sebelumnya, hanya analisis ekstraktif yang digunakan dan dibandingkan dengan hasil Microsoft Word. [6].

B. Abstraksi

Peringkasan Abstraksi 12 Teknik abstraksi menggunakan keterampilan bahasa untuk mempelajari dan menafsirkan teks dari dokumen resume. Buat ringkasan teks dengan menambahkan kalimat baru yang mewakili esensi teks asli dalam bentuk yang berbeda dari kalimat yang ada dalam teks asli. [9].

3.3. Pra Proses

Preprocessing adalah tahap dimana teks disiapkan sebagai data yang akan diproses pada langkah selanjutnya. Preprocessing teks yang digunakan dalam penelitian ini menggunakan proses pembentukan kalimat, case folding, tokenization, dan stopwords.. Tahap preprocessing dapat dilihat pada Gambar 2 berikut:



Gambar 2. Tahapan Pra Proses

A. Segmentasi

Segmentasi yaitu proses membentuk string teks dokumen yang Panjang menjadi kumpulan kalimat-kalimat atau paragraf - paragraf dengan cara dipisahkan. Pemisahan teks menjadi kalimat-kalimat, atau paragraph - paragraf menghasilkan korespondensi kata yang lebih baik. Penelitian sebelumnya menggunakan segmentasi untuk memotong bahasa 'Tigrinya', sesuai aturan yang telah ditentukan untuk menaikkan kualitas data dan membuat token yang baik [10]. Penelitian terdahulu lainnya, melakukan proses mengubah dokumen menjadi kalimat-kalimat, terlebih dahulu mengubah file berformat .pdf ke string, setelah itu digunakan fungsi split(), dengan tanda titik ".", tanda tanya "?", dan tanda seru "!" sebagai delimiter untuk memotong string dokumen [11].

B. Case Folding

Case folding adalah proses konversi huruf dalam teks dimana semua huruf dalam teks diubah ke dalam case yang sama. Pada penelitian terdahulu penggunaan case folding digunakan pada tahap pra proses untuk membersihkan data teks, merubah huruf besar menjadi huruf kecil, sehingga tidak ada redundansi data [12].

C. Tokenizing

Tokenizing adalah proses pemisahan teks yang ada pada data makalah menjadi potongan - potongan kecil

yang disebut token. Pada penelitian sebelumnya tokenizing digunakan untuk pembagian hak cipta memungkinkan pembagian keuntungan otomatis dan efisien [13].

D. Stopwords

Daftar stopword atau stopword bukanlah kata deskriptif (non-kritis) yang dapat dibuang dengan menggunakan pendekatan back-of-word.. Penelitian A method to generate text summary by accounting pronoun frequency for keywords weightage computation menggunakan stopwords untuk menghilangkan kata – kata tidak penting, dan menghasilkan kata – kata penting [14].

3.4. Term Frequency – Inverse Document Frequency

Term Frequency - Inverse Document Frequency (TF-IDF) merupakan metode ekstraksi kalimat dengan cara memberikan nilai atau bobot pada kata. Pembobotan kata bertujuan untuk mencerminkan betapa pentingnya sebuah kata dalam sebuah kumpulan dokumen atau korpus [15].

3.5. Maximum Marginal Relevance

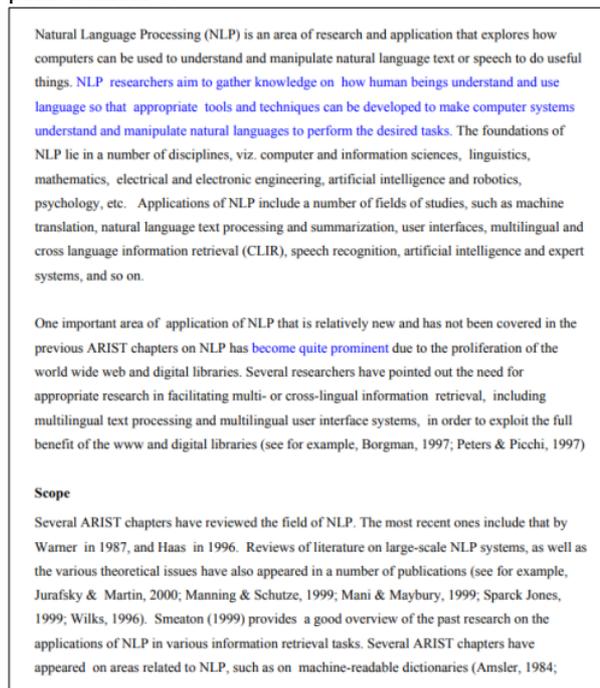
Algoritma Maksimum Margin Relevansi (MMR) adalah salah satu metode peringkasan yang digunakan untuk meringkas dokumen individu atau beberapa dokumen. MMR merangkum dokumen dengan menghitung derajat kemiripan (similarity) antar bagian teks. Proposal dengan MMR tertinggi dipilih secara iteratif dari resume sampai nilai yang telah ditentukan tercapai. Penelitian-penelitian sebelumnya menggunakan metode maksimal marginalrelevance digunakan untuk menghitung bobot ranking dari setiap proposal. [16].

4. PERANCANGAN SISTEM PERINGKASAN

Pada penelitian ini proses peringkasan dilakukan menggunakan Term Frequency - Inverse Document Frequency dan maximum marginal relevance. Data yang akan menjadi masukan berupa data jurnal dan proseding yang sesuai kriteria yang disimpan dalam format (.pdf). Data yang dimasukkan akan dilakukan preprocessing, yaitu proses segmentasi dari teks secara keseluruhan akan dipisah menjadi bentuk kumpulan kalimat-kalimat. Hasil segmentasi akan tokenasi, pemisahan dari kumpulan kalimat menjadi kumpulan kata. Praproses terakhir melakukan stopwords removing, penghapusan karakter-karakter, dan kata-kata tidak perlu seperti kata sambung. Tahap proses, data dari tahap preproses akan masuk ke TF-IDF, pertama penghitung TF, perhitungan IDF, dan perhitungan TFIDF. Selanjutnya masuk ke MMR, data setelah proses TF-IDF, menghitung similarity kalimat dengan koleksi, menghitung nilai maksimal dari setiap kalimat, 20 terakhir mengeleminasi kalimat yang tidak sesuai kriteria, penggabungan kalimat sesuai kriteria menjadi ringkan.

4.1. Perolehan Data

Pada penelitian ini hanya terdapat data . Data makalah ilmiah, jurnal dan proseding didapatkan dari banyak web resmi penyedia makalah ilmiah seperti Google Scholar, Lembaga Ilmu Pengetahuan Indonesia (LIPI), Perpustakaan Nasional Republik Indonesia (Perpusnas), Directory of Open Access Journals (DOAJ), Citeseer, Organization of Scientific Research Journals dan sebagainya. Contoh makalah dapat dilihat pada Gambar 3.



Gambar 3. Contoh Makalah

4.2. Pra Proses

Pada tahap praproses, data makalah diolah untuk mendapatkan data masukan yang baik. Tahapan praproses dilakukan dengan empat tahap yaitu segmentasi, case folding, tokenezing, dan stopwords removal.

A. Segmentasi

Segmentasi merupakan proses pemotongan data menjadi beberapa bagian. Pemotongan data dalam makalah ilmiah akan dipisah menjadi kumpulan kalimat. Tujuan dari proses segmentasi adalah menadapatkan kalimat. Dokumen makalah ilmiah, kalimat bisa dipisahkan dengan tanda baca seperti tanda titik “.”, tanda tanya “?”, dan tanda seru “!”. Segemntasi pada dokumen makalah dapat dilihat dilihat pada Tabel 1.

Tabel 1. Hasil Segmentasi

No	Kalimat
1	Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things
2	NLP researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to

perform the desired tasks

3 The foundations of NLP lie in a number of disciplines, viz

4 computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology, etc

5 Applications of NLP include a number of fields of studies, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval (CLIR), speech recognition, artificial intelligence and expert systems, and so on

6 One important area of application of NLP that is relatively new and has not been covered in the previous ARIST chapters on NLP has become quite prominent due to the proliferation of the world wide web and digital libraries

7 Several researchers have pointed out the need for appropriate research in facilitating multi- or cross-lingual information retrieval, including multilingual text processing and multilingual user interface systems, in order to exploit the full benefit of the www and digital libraries (see for example, Borgman, 1997; Peters & Picchi, 1997)

8 Several ARIST chapters have reviewed the field of NLP. The most recent ones include that by Warner in 1987, and Haas in 1996.

9 Reviews of literature on large-scale NLP systems, as well as the various theoretical issues have also appeared in a number of publications (see for example, Jurafsky & Martin, 2000; Manning & Schutze, 1999; Mami & Maybury, 1999; Sparck Jones, 1999; Wilks, 1996)

10 Smeaton (1999) provides a good overview of the past research on the applications of NLP in various information retrieval tasks. Several ARIST chapters have appeared on areas related to NLP, such as on machine-readable dictionaries (Amsler, 1984;

B. Case Folding

Data teks yang terpotong menjadi kalimat direduksi menjadi case pada langkah ini, proses konversi teks ke format standar atau huruf kecil dan menghilangkan angka, tanda baca dan simbol dianggap sebagai pembatas, sehingga sistem hanya menerima karakter. dan simbol.. Hasil dari case folding bisa dilihat pada Tabel 2.

Tabel 2. Hasil Case Folding

No	Kalimat
1	natural language processing nlp is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things
2	nlp researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to perform the desired tasks
3	the foundations of nlp lie in a number of disciplines viz
4	computer and information sciences linguistics mathematics electrical and electronic engineering artificial intelligence and robotics psychology etc
5	applications of nlp include a number of fields of studies such as machine translation natural language text processing and summarization user interfaces multilingual and cross language information retrieval clir speech recognition artificial intelligence and expert systems and so on
6	one important area of application of nlp that is relatively new and has not been covered in the previous arist chapters on nlp has become quite prominent due to the proliferation of the world wide web and digital libraries
7	several researchers have pointed out the need for appropriate research in facilitating multi or cross lingual information retrieval including multilingual text processing and multilingual user interface systems in

order to exploit the full benefit of the www and digital libraries see for example borgman 1997 peters picchi

8 several arist chapters have reviewed the field of nlp the most recent ones include that by warner in and has in

9 reviews of literature on large-scale nlp systems, as well as the various theoretical issues have also appeared in a number of publications see for example jurafsky martin manning schutze mani maybury sparck jones wilks

10 smeaton provides a good overview of the past research on the applications of nlp in various information retrieval tasks. several arist chapters have appeared on areas related to nlp such as on machine readable dictionaries amsler

C. Tokenizing

Kalimat yang dihasilkan dari proses segmentasi akan dipotong menjadi beberapa token/bagian. Token/bagian itu bisa berupa frasa atau kata. Pada proses ini akan dipotong menjadi kumpulan kata. Pemisahan dilakukan dengan memanfaatkan delimeter spasi “ “. Hasil dari dari proses word tokenezing dilihat pada dilihat pada Tabel 3.

Tabel 3. Hasil Tokenizing

Kata	Kata	Kata	Kata
natural	language	processing	nlp
is	an	area	of
research	and	application	that
explores	how	computers	can
be	used	to	understand
and	manipulate	natural	language
text	or	speech	to
do	useful	things	nlp
researchers	aim	to	gather
knowledge	on	how	human
beings	understand	and	use
language	so	that	appropriate
tools	and	techniques	can
be	developed	to	make
computer	systems	understand	and
manipulate	natural	languages	to
perform	the	desired	tasks

D. Stopwords

Stopwords removal merupakan proses penghapusan kata-kata yang dianggap tidak penting. Tahap ini digunakan untuk mengurangi proses peringkasan. Hasil dari stopwords removal terlihat pada Tabel 4.

Tabel 4. Hasil Stopwords

Kata	Kata	Kata	Kata
natural	language	processing	nlp
area	manipulate	natural	language
research	application	explores	computers
used	understand	text	speech
useful	things	nlp	tasks
researchers	aim	gather	desired
knowledge	human	language	appropriate
beings	understand	use	perform
tools	and	techniques	languages
developed	make	manipulate	natural
computer	systems	understand	and

E. Term Frequency – Inverse Document Frequency

Pembobotan otomatis menentukan jumlah kemunculan suatu kata (term) dalam suatu dokumen

(frekuensi kemunculan suatu istilah) dan jumlah kemunculan dalam kumpulan dokumen (frekuensi kemunculan (inverse Occurency) suatu kata dalam suatu dokumen) . Arti sebuah kata lebih besar ketika sering muncul dalam dokumen dan kurang penting ketika muncul di banyak dokumen.. Terdapat beberapa Tahapan untuk mengetahui bobot diantaranya adalah :

1. Menghitung Term Frequency (TF).

Tabel 5. Term Frequency

Term	Kalimat 1	Kalimat 2	Kalimat 3	Kalimat 4
Term1	0	0	1	0
Term2	0	1	0	0
Term3	1	0	0	2
Term4	0	3	0	2

2. Menghitung Document Frequency (DF).

Tabel 6. Document Frequency

Term	Kalimat Frequency
Term1	1
Term2	1
Term3	2
Term4	5

3. Menghitung Invers Document Frequency.

Tabel 7. Invers Document Frequency

Term	Kalimat Frequency
Term1	$\text{Log}(4/1) = 0,60205$
Term2	$\text{Log}(4/1) = 0,60205$
Term3	$\text{Log}(4/2) = 0,30102$
Term4	$\text{Log}(4/3) = 0,1249$

4. Menghitung Term Frequency - invers Dokument Frequency (TF-IDF).

Tabel 8. Term Frequency – invers document frequency

tf	idf	Kalimat 1	Kalimat 2	Kalimat 3	Kalimat 4
1	$\text{Log}(4/1) = 0,60205$	0	0,60205	0,60205	0
1	$\text{Log}(4/1) = 0,60205$	0	0,60205	0	0
2	$\text{Log}(4/2) = 0,30102$	0,30102	0	0	0,60204
5	$\text{Log}(4/3) = 0,1249$	0,30102	0,52627	0,60205	0,42592

4.3. Maximum Marginal Relevance

Maximum marginal relevance (MMR) merupakan salah satu metode ekstraksi ringkasan (extractive summary) yang digunakan untuk meringkas dokumen tunggal atau multi dokumen. Kalimat dengan nilai MMR tertinggi dari setiap perhitungan iterasi akan diambil, kemudian dipilih untuk dijadikan sebagai ringkasan. Iterasi akan berhenti ketika hasil MMR maksimum sama dengan 0 atau 1. Untuk peringkasan small dokumen, seperti pada berita (news), menggunakan nilai parameter $\lambda=0.7$.

1. Perhitungan Iterasi ke- 1

Nilai $Sim1 (S_i, S')$ adalah kosong (0).

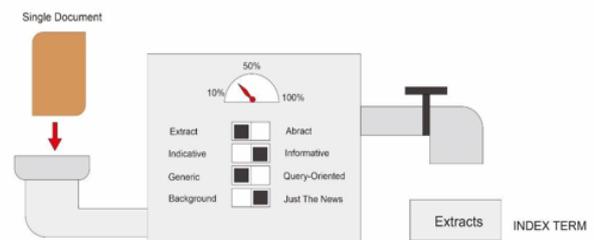
$$MMR (1) = 0,7 * Sim1 (S1,Q) - (1 - 0,7) * \max Sim2 (S1, S')$$

$$MMR (2) = 0,7 * Sim1 (S2,Q) - (1 - 0,7) * \max Sim2 (S2, S')$$

$$MMR (3) = 0,7 * Sim1 (S3,Q) - (1 - 0,7) * \max Sim2 (S3, S')$$

$$MMR (4) = 0,7 * Sim1 (S4,Q) - (1 - 0,7) * \max Sim2 (S4, S')$$

Perhitungan MMR akan berhenti sesuai dengan dengan yang ditentukan, setiap iterasi akan mendapatkan nilai dari setiap kalimat yang akan menjadi ringkasan. Similitas yang dihasilkan pada setiap kalimat menggunakan MMR akan dimaukan ke dalam array sesuai dengan dengan urutan kalimat yang terdapat dalam dokumen makalah. Nilai similaritas kalimat menjadi acuan pembuatan ringkasan. Setelah penghitungan similaritas, masuk pada pengurutan nilai similaritas kalimat, menggunakan Insert Sort, nilai similaritas yang ada pada array 29 indeks ke 1 akan dibandingkan nilai terbesar dengan indeks ke 2, hasilnya akan dimasukan ke array baru penampung nilai similaritas yang telah tersusun dari yang terbesar ke terkecil, selanjutnya indeks ke 3 dibandingkan dengan nilai similaritas yang telah terisi di array baru jika nilai similitas pembanding lebih besar dari yang dibandingkan maka akan diletakan di indeks didepan yang dibandingkan dan menggeser indeks yang ada dibelakangnya jika lebih kecil maka akan dibandingkan dengan dengan nilai atau elemen pembanding selanjutnya pada array yang telah tersusun, jika sampai indeks terakhir pembanding masih lebih kecil dari yang dibandingkan maka diletakan pada indeks selanjutnya pada array nilai similaritas yang telah terurut, hal tersebut diulangi sampai semua similaritas yang terdapat dalam array belum terurut dibandingkan semuanya. Selanjutnya melakukan perhitungan perkalian antara koefisien ringkasan(persentase kompresi/peringkasan) dengan jumlah elemen array nilai similaritas telah terurut, untuk mendapatkan nilai jumlah kalimat (Indeks term) yang akan disatukan menjadi ringkasan atau untuk menyeleksi kriteia kalimat untuk dibuat ringkasan. Selanjutnya indeks term digunakan untuk mengambil kumpulan kalimat dari indeks awal sampai sampai indeks peringkasan dan menggabungkan kalimat-kalimat tersebut menjadi sebuah sebuah ringkasan. Mekanisme peringkasan terlihat pada Gambar 4.



Gambar 4. Mesin Peringkasan

5. KESIMPULAN

Penelitian ini menghasilkan perangkat lunak peringkasan makalah otomatis dengan Maximum Marginal Relevance dapat melakukan peringkasan terhadap makalah. Hasil akhir dari penelitian ini adalah data teks yang telah diringkas dari data teks semula. Hasil penelitian menggunakan Rouge Score dengan membandingkan peringkasan dengan MMR dengan peringkasan esktraktif. peringkasan ekstraktif mendapatkan score Rouge-1 terbaik dengan akurasi 67% dikarenakan penggunaan praproses TF-IDF, dan fokus hanya terhadap token. Sedangkan peringkasan menggunakan MMR mendapatkan nilai akurasi terbaik dari penilain Rouge-2 akurasi 60%, karena penggunaan MMR tidak mengubah susuna kata pada kalimat, Rouge-L akurasi 68%, dan Rouge-W akurasi 36%. Peringkasan otomatis menggunakan Maximum Marginal Relevance lebih baik dari segi menghasilkan lebih sedikit dari tumpang tindih kalimat, menghasilkan struktur ringkasan lebih mirip dengan abstrak dari makalah. Seluruh Pengujian memeberikan hasil possitif dengan kesimpulan bahwa peringkasan makalah ilmiah otomatis dengan maximum marginal relevance untuk melakukan peringkasan makalah ilmiah, menghasilkan hasil rangkuman dengan hasil pengujian terbaik pada pengujian Rouge-1 dengan akurasi 64%, Rouge-2 dengan akurasi 60%, Rouge-L dengan akurasi 68%, dan Rouge-W dengan akurasi 36%.

DAFTAR PUSTAKA

- [1] R. Flüthmann and R. Ostermann, "Ein EDV-standardreport für die MDK-Prüfung in der stationären Altenhilfe: Den ersten Einblick verschaffen.," *Pflege Z.*, vol. 62, no. 6, pp. 337–341, 2009.
- [2] V. Gupta and G. S. Lehal, "A Survey of Text Summarization Extractive techniques," *J. Emerg. Technol. Web Intell.*, vol. 2, no. 3, pp. 258–268, 2010, doi: 10.4304/jetwi.2.3.258-268.
- [3] R. Ilyas and F. Umbara, "Peringkasan Otomatis Dengan Ekstraksi Informasi Untuk Dokumen Berita Ter- cluster," in *ANNUAL RESEARCH SEMINAR 2016*, 2016, vol. 2, no. 1, pp. 405–408.
- [4] C. J. Jiang, H. Peng, Q. L. Ma, and J. C. Chen, "Automatic summarization for Chinese text based on combined words recognition and paragraph clustering," *3rd Int. Symp. Intell. Inf. Technol. Secur. Informatics, IITSI 2010*, pp. 591–594, 2010, doi: 10.1109/IITSI.2010.15.
- [5] S. Setiadi B, "Peringkasan Kumpulan Berita Secara Otomatis Menggunakan Metode Maximum Marginal Relevance," in *Prosiding Seminar Nasional Teknologi dan Informatika*, 2017, vol. 123, pp. 153–160, doi: 10.2298/PAN0903301G.
- [6] J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," in *2019 International Conference on Data Science and Communication, IconDSC 2019*, 2019, pp. 1–3, doi: 10.1109/IconDSC.2019.8817040.
- [7] D. R. Radev and G. Erkan, "LexRank : Graph-based Centrality as Saliency in Text Summarization," *J. Artif. Intell. Res.*, vol. 22, no. 1, pp. 457–479, 2004, [Online]. Available: <https://arxiv.org/abs/1109.2128>.
- [8] V. Dalal and L. Malik, "A survey of extractive and abstractive text summarization techniques," *Int. Conf. Emerg. Trends Eng. Technol. ICETET*, pp. 109–110, 2013, doi: 10.1109/ICETET.2013.31.
- [9] J. Chi, L. Li, and Z. Huang, "Mining the Usage of Summary Oriented Features in Abstractive Summarization," *Proc. 2018 5th IEEE Int. Conf. Cloud Comput. Intell. Syst. CCIS 2018*, pp. 789–793, 2018, doi: 10.1109/CCIS.2018.8691296.
- [10] Y. Tedla and K. Yamamoto, "The effect of shallow segmentation on English-Tigrinya statistical machine translation," *Proc. 2016 Int. Conf. Asian Lang. Process. IALP 2016*, pp. 79–82, 2017, doi: 10.1109/IALP.2016.7875939.
- [11] U. Nuha and N. Rochmawati, "Klasifikasi Kesahihan Hadits Berdasarkan Perawi Hadits Menggunakan Principal Component Analysis (PCA) dan Backpropagation Neural Network (BPNN)," (*Journal Informatics Comput. Sci.*, vol. 01, no. 02, pp. 138–143, 2019.
- [12] R. A. Pane, M. S. Mubarak, N. S. Huda, and Adiwijaya, "A multi-label classification on topics of Quranic verses in English translation using multinomial naive bayes," *2018 6th Int. Conf. Inf. Commun. Technol. ICoICT 2018*, vol. 0, no. 1, pp. 481–484, 2018, doi: 10.1109/ICoICT.2018.8528777.
- [13] T. Ishida *et al.*, "Tokenizing and managing the copyrights of digital content on blockchains," *2019 IEEE 8th Glob. Conf. Consum. Electron. GCCE 2019*, pp. 170–172, 2019, doi: 10.1109/GCCE46687.2019.9015595.
- [14] S. Urolagin, J. Nayak, and L. Satish, "A method to generate text summary by accounting pronoun frequency for keywords weightage computation," *Proc. 2017 Int. Conf. Eng. Technol. ICET 2017*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.1109/ICEngTechnol.2017.8308170.
- [15] K. E. Dewi and E. Widiastuti, Nelly Indriani Rainarli, "Evaluasi Sentence Extraction pada Peringkasan Dokumen Otomatis," *Pros. Semin. Nas. Inform. dan Apl.*, no. September, p. A-8 sd. A-12, 2017.
- [16] U. Shasha Xie , Yang Liu The University of Texas at Dallas , Richardson , TX, "Using Corpus And Knowledge-Based Similarity Measure In Maximum Marginal Relevance For Meeting Summarization," *Computer (Long. Beach. Calif.)*, vol. 2, no. 2, pp. 4985–4988, 2008.