

**PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN
MENGUNAKAN ALGORITMA TEXTRANK**

SKRIPSI



disusun oleh

Fia Amara Sandra Hardiana

18.11.2491

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN
MENGUNAKAN ALGORITMA TEXTRANK**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Fia Amara Sandra Hardiana

18.11.2491

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN MENGUNAKAN ALGORITMA TEXTRANK

yang dipersiapkan dan disusun oleh

Fia Amara Sandra Hardiana

18.11.2491

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Oktober 2021

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom.
NIK. 190302108

PENGESAHAN
SKRIPSI
PENDEKATAN EKSTRAKTIF PADA SUMMARIZE DOKUMEN
MENGGUNAKAN ALGORITMA TEXTRANK

yang dipersiapkan dan disusun oleh

Fia Amara Sandra Hardiana

18.11.2491

telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Februari 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arif Dwi Laksito, M.Kom
NIK. 190302150

Muhammad Rudyanto Arief, M.T
NIK. 190302098

Mardhiya Hayaty, S.T., M.Kom
NIK. 190302108

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal ... April 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 23 Maret 2022



Fia Amara Sandra Hardiana

NIM. 18.11.2491

MOTTO

"Hatiku tenang karena mengetahui bahwa apa yang melewatkanmu tidak akan pernah menjadi takdirku. dan apa yang ditakdirkan untukku, takkan melewatkanmu."

(Umar Bin Khattab)

"Apabila sesuatu yang kau senangi tidak terjadi maka senangilah apa yang terjadi."

(Ali bin Abi Thalib)



PERSEMBAHAN

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan lancar. Selama pengerjaan skripsi ini banyak sekali bantuan, dukungan, dan do'a dari berbagai pihak, sehingga penulis sampaikan rasa terimakasih sedalam-dalamnya kepada:

1. Kedua orang tua, kakak, dan saudara-saudara yang telah memberikan do'a dan dukungan kepada penulis secara moril maupun materil sehingga skripsi ini dapat selesai dengan baik.
2. Dosen pembimbing, Ibu Mardhiya Hayaty, S.T., M.Kom. yang telah membimbing dalam penyusunan skripsi ini hingga selesai.
3. Seluruh dosen dan staff Universitas Amikom Yogyakarta yang telah membantu dan membimbing penulis selama proses perkuliahan.
4. Sahabat-sahabat penulis Evanisa Ananda, Zulfikar Murakabiman, Rizka Dewi Anggraini yang telah memberikan pengalaman dan membantu menyelesaikan permasalahan.
5. Teman-teman Informatika angkatan 2018 khususnya kelas IF-10 yang senantiasa memberi motivasi dan berjuang bersama selama menjadi mahasiswa.
6. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak bisa penulis sebutkan satu-persatu.

KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT, karena rahmat dan hidayah-Nya sehingga skripsi yang berjudul “Pendekatan Ekstraktif pada Summarize Dokumen Menggunakan Algoritma *TextRank*” ini dapat selesai dengan baik sebagai salah syarat untuk dapat menempuh ujian sarjana pada Fakultas Ilmu Komputer (FIK) program studi Informatika di Universitas Amikom Yogyakarta.

Pengerjaan skripsi ini tidak lepas dari bantuan beberapa pihak. Oleh karena itu, penulis ingin menyampaikan rasa hormat dan terimakasih kepada:

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Windha Mega Pradnya Duhita, M.Kom. selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
3. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku dosen pembimbing yang telah membantu dan membimbing penulis dengan saran dan waktunya.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan sehingga saran dan kritik yang membangun sangat penulis harapkan dan penulis berharap semoga skripsi ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi.

Yogyakarta,

Fia Amara Sandra Hardiana

DAFTAR ISI

JUDUL	I
PERSETUJUAN.....	II
PENGESAHAN.....	III
PERNYATAAN	IV
MOTTO.....	V
PERSEMBAHAN.....	VI
KATA PENGANTAR	VII
DAFTAR ISI.....	VIII
DAFTAR TABEL	X
DAFTAR GAMBAR.....	XI
INTISARI.....	XII
<i>ABSTRACT</i>	XIII
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	3
1.3 BATASAN MASALAH	3
1.4 MAKSUD DAN TUJUAN PENELITIAN	4
1.5 MANFAAT PENELITIAN	4
1.6 METODE PENELITIAN	4
1.7 SISTEMATIKA PENULISAN.....	5
BAB II LANDASAN TEORI	7
2.1 KAJIAN PUSTAKA.....	7
2.2 DASAR TEORI.....	9
2.2.1 <i>Artikel Berita</i>	9

2.2.2	<i>Peringkasan Teks</i>	10
2.2.3	<i>Preprocessing</i>	12
2.2.4	<i>Ekstraksi Fitur (Feature Extraction)</i>	13
2.2.5	<i>Algoritma TextRank</i>	14
2.2.6	<i>Evaluasi Ringkasan</i>	18
BAB III METODE PENELITIAN		20
3.1	ALUR PENELITIAN	20
3.2	DATASET	21
3.3	<i>PREPROCESSING</i>	22
3.4	EKSTRAKSI FITUR (<i>FEATURE EXTRACTION</i>).....	22
3.5	<i>PROCESSING</i>	22
3.6	EVALUASI HASIL RINGKASAN	23
BAB IV HASIL DAN PEMBAHASAN		24
4.1	HASIL	24
4.2	PEMBAHASAN	24
4.2.1	<i>Dataset</i>	24
4.2.2	<i>Preprocessing</i>	26
4.2.3	<i>Ekstraksi Fitur (Feature Extraction)</i>	28
4.2.4	<i>Processing</i>	28
4.2.5	<i>Evaluasi</i>	39
BAB V PENUTUP		42
5.1	KESIMPULAN	42
5.2	SARAN.....	42
DAFTAR PUSTAKA		43

DAFTAR TABEL

Tabel 2. 1 Penelitian terdahulu	8
Tabel 4. 1 Skor ROUGE pada dataset Liputan6.....	24
Tabel 4. 2 Hasil segmentation	26
Tabel 4. 3 Hasil case folding	27
Tabel 4. 4 Hasil stopwords	27
Tabel 4. 5 Hasil vektor 5 dimensi.....	29
Tabel 4. 6 Hasil cosine similarity	31
Tabel 4. 7 Inisialisasi awal bobot PageRank	34
Tabel 4. 8 Hasil perhitungan bobot PageRank	37
Tabel 4. 9 Hasil ranking bobot PageRank.....	38
Tabel 4. 10 Hasil ringkasan dokumen.....	38
Tabel 4. 11 Perhitungan kata yang sama dari ringkasan.....	39
Tabel 4. 12 Perhitungan evaluasi ringkasan dengan vektor 5 dimensi	40
Tabel 4. 13 Perhitungan kata yang sama dari ringkasan.....	40
Tabel 4. 14 Perhitungan evaluasi ringkasan dengan vektor 300 dimensi	41

DAFTAR GAMBAR

Gambar 2. 1	Arsitektur peringkasan ekstraktif.....	11
Gambar 2. 2	Arsitektur peringkasan abstraktif.....	12
Gambar 2. 3	Graf sentence extraction TextRank	17
Gambar 3. 1	Diagram proses	21
Gambar 4. 1	Bentuk dataset Liputan6.....	25
Gambar 4. 2	Bentuk dataset setelah pembersihan	25
Gambar 4. 3	Dataset yang siap digunakan	26
Gambar 4. 4	Kata terdekat dari lelaki	28
Gambar 4. 5	Vektor kalimat	28
Gambar 4. 6	Similarity matrix.....	32
Gambar 4. 7	Graf similarity	33
Gambar 4. 8	Ilustrasi graf.....	34
Gambar 4. 9	Bobot PageRank konvergen	38
Gambar 4. 10	Hasil ranking bobot PageRank konvergen	39

INTISARI

Perkembangan teknologi dan informasi yang pesat mengakibatkan informasi yang tersedia baik online maupun offline semakin bertambah salah satunya adalah artikel berita. Dari banyaknya kalimat pada artikel berita, hanya beberapa informasi penting yang merupakan representasi pokok pikiran dari penulis. Artikel berita bersifat *real time* dan *up to date* yang mengakibatkan pembaca harus meluangkan waktu dalam membaca dan mengikuti perkembangan berita, sehingga dibutuhkan sebuah peringkasan teks otomatis. Peringkasan teks otomatis (*Automatic Text Summarization*) merupakan peringkasan teks yang dilakukan oleh komputer untuk membuat sebuah artikel menjadi lebih sederhana dengan mengambil kesimpulan tanpa menghilangkan makna yang ada.

Salah satu algoritma yang dapat digunakan untuk meringkas teks otomatis adalah algoritma *TextRank*. *TextRank* merupakan metode perangkangan kalimat berbasis *graph* dengan menentukan nilai tertinggi dari sebuah kalimat. Setiap kalimat akan dianggap sebagai sebuah *vertex* dan akan dihubungkan berdasarkan nilai *similarity* menggunakan *cosine similarity*. Pembobotan pada setiap *vertex* menggunakan perhitungan *PageRank* dan ringkasan yang dihasilkan mengambil satu kalimat teratas.

Hasil evaluasi ringkasan artikel berita berbahasa Indonesia dengan membandingkan ringkasan mesin yang diperoleh menggunakan algoritma *TextRank* dengan ringkasan manusia pada dataset Liputan6 sebanyak 193.883 menghasilkan nilai rata-rata *f-measure* pada ROUGE-1 sebesar 27.82%, ROUGE-2 sebesar 14.22%, dan ROUGE-L sebesar 23.13%.

Kata Kunci: Peringkasan Teks, *Extractive Summarization*, *TextRank*, ROUGE

ABSTRACT

The rapid development of technology and information has resulted in increasing information available both online and offline, one of which is news articles. Of the many sentences in the news article, only a few important information represents the main idea of the author. News articles are real time and up to date which causes readers to have to spend time reading and following news developments, so an automatic text summary is needed. Automatic text summarization is a text summarization carried out by a computer to make an article simpler by drawing conclusions without losing the existing meaning.

One of the algorithms that can be used to summarize text automatically is the TextRank algorithm. TextRank is a graph-based sentence ranking method by determining the highest value of a sentence. Each sentence will be considered as a vertex and will be linked based on the similarity value using cosine similarity. The weighting on each vertex uses a PageRank calculation and the resulting summary takes the top one sentence.

The results of the evaluation of the summary of Indonesian news articles by comparing the machine summary obtained using the TextRank algorithm with the human summary on the Liputan6 dataset of 193,883 resulted in an average f-measure value of 27.82% for ROUGE-1, 14.22% for ROUGE-2, and ROUGE-L is 23.13%.

Keyword: Text Summarization, Extractive Summarization, TextRank, ROUGE

