

TESIS

**PERINGKASAN TEKS BAHASA INDONESIA
SECARA OTOMATIS MENGGUNAKAN METODE
SENTENCE SCORING DAN DECISION TREE**



PERIANTU MARHENDRI SABUNA
No. Mhs.: 155302367/PS/MTF

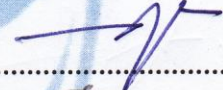
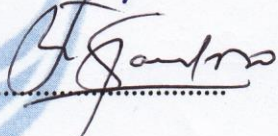
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA
PROGRAM PASCASARJANA
UNIVERSITAS ATMA JAYA YOGYAKARTA
2017



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : PERIANTU MARHENDRI SABUNA
Nomor Mahasiswa : 155302367/PS/MTF
Konsentrasi : Enterprise Information Sistem
Judul Tesis : PERINGKASAN TEKS BAHASA INDONESIA
SECARA OTOMATIS MENGGUNAKAN METODE
SENTENCE SCORING DAN DECISION TREE

Nama Pembimbing	Tanggal	Tanda tangan
Ir. A. Djoko Budiyanto, M. Eng., Ph. D	21-7-2017	
Dr. Ir. Alb. Joko Santoso, MT	21-7-2017	



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : PERIANTU MARHENDRI SABUNA
Nomor Mahasiswa : 155302367/PS/MTF
Konsentrasi : Enterprise Information Sistem
Judul Tesis : PERINGKASAN TEKS BAHASA INDONESIA
SECARA OTOMATIS MENGGUNAKAN METODE
SENTENCE SCORING DAN DECISION TREE

Nama Penguji

Tanggal

Tanda tangan

Ir. A. Djoko Budiyo, M. Eng., Ph. D
(Ketua)

21-7-2017

Dr. Ir. Alb. Joko Santoso, MT.
(Sekretaris)

21-7-2017

Dr. Pranowo, ST. MT.
(Anggota)

21/7/2017



Ketua Program Studi

Prof. Ir. Suyoto, M. Sc., Ph. D

PROGRAM
PASCASARJANA



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PERNYATAAN

Bersamaan dengan penelitian ini, maka saya yang bertanda tangan di bawah ini :

Nama : PERIANTU MARHENDRI SABUNA
Nomor Mahasiswa : 155302367/PS/MTF
Konsentrasi : Enterprise Information Sistem
Judul Tesis : PERINGKASAN TEKS BAHASA INDONESIA
SECARA OTOMATIS MENGGUNAKAN METODE
SENTENCE SCORING DAN DECISION TREE

Menyatakan bahwa penelitian ini adalah hasil pemikiran sendiri dan bukan duplikasi dari karya tulis yang telah ada sebelumnya. Karya tulis yang telah ada sebelumnya dijadikan acuan oleh penulis guna melengkapi penelitian ini dan dinyatakan secara tertulis dalam penulisan acuan dan daftar pustaka.

Demikian pernyataan ini dibuat untuk digunakan sebagaimana mestinya.

Yogyakarta, Juli 2017

PERIANTU MARHENDRI SABUNA

INTISARI

Peringkasan teks adalah proses pemampatan teks sumber ke dalam versi lebih pendek namun tetap mempertahankan informasi pokok yang terkandung didalamnya. Memahami isi dokumen berita melalui ringkasan teks memerlukan waktu yang lebih singkat dibandingkan membaca seluruh isi dokumen, sehingga ringkasan teks menjadi sangat penting. Dengan adanya ringkasan, diharapkan pembaca dapat dengan cepat dan mudah memahami makna sebuah teks tanpa harus membaca keseluruhan teks. Oleh karena itu, maka perlu menggunakan metode untuk memahami, mengklasifikasikan dan menyajikan semua informasi dengan cara yang jelas dan ringkas, memungkinkan pembaca untuk menghemat waktu dan sumber daya.

Dalam penelitian ini akan mengkombinasikan metode *sentence scoring* dan *decision tree* untuk peringkasan teks secara otomatis dalam bahasa Indonesia. Algoritma *decision tree* digunakan untuk memilih kalimat mana saja yang akan masuk dalam ringkasan sistem. Untuk menghasilkan rule-rule dalam pohon keputusan, akan digunakan 50 teks berita sebagai data training. Model yang dihasilkan pada tahap training diterapkan pada proses seleksi kalimat yang akan masuk dalam ringkasan sistem. Hasil penelitian menunjukkan pengukuran nilai f-measure tertinggi adalah 0,80 dan rata-rata 0,58. Berdasarkan hasil tersebut dapat disimpulkan bahwa hasil peringkasan dokumen menggunakan *sentence scoring* dan *decision tree* menunjukkan nilai akurasi yang cukup baik untuk dokumen teks berita.

Kata kunci : peringkasan teks, *sentence scoring*, *decision tree*

ABSTRACT

Text summarization is a process of compressing a text from the source to be a shorter version, but the version still contains the main information there. By reading the summary, the readers might be easy and fast to understand the contents instead of reading all the text. Because of that, it needs a method to understand, clarify, and present the whole information needed clearly and succinctly in the summary. So, it allows the readers save the time and energy.

This research combining sentence scoring and decision tree method for automatic text summarization in Indonesian language. It uses the decision tree algorithm to choose which of sentences will be selected in summarization system. To produce the rules for decision tree, it uses 50 news texts as the training data. The produced-model from the training stage will be implemented for sentence selection process to the summarization system. The result shows the highest f-measure score is 0, 80 and the average is 0, 58. Based on this, it concludes that the result of document summarization using sentence scoring and decision tree shows a better accuracy score for news text document.

Keywords : *text summarization, sentence scoring, decision tree*

LEMBAR PERSEMBAHAN

“Untuk menjadi yang terbaik,
berikanlah usaha terbaik mu....”

“Kesuksesan adalah saat dimana seseorang berhasil
memaksimalkan potensi dirinya....”

“Akuilah Dia dalam segala laku-mu, maka Ia akan meluruskan
jalan-mu....”

- Amsal 3 : 6

Karya ini kupersembahkan untuk :

Tuhan Yesus sebagai pemberi hikmah

Ayah dan Ibu Tercinta
(Bapa Paul & mama Ros)

Kekasih tersayang Vara Sikas

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yesus Kristus yang telah memberikan rahmat dan anugerah-Nya sehingga penulis dapat menyelesaikan penelitian Tesis dengan baik dan dapat menghasilkan tulisan yang dapat berguna bagi banyak orang yang membacanya.

Penulisan tesis ini dimaksudkan untuk memenuhi sebagian persyaratan mencapai derajat Magister Teknik Informatika dari Program Pascasarjana Universitas Atma Jaya Yogyakarta sekaligus menerapkan ilmu yang telah diperoleh selama studi. Penulis menyadari bahwa selesainya pembuatan tesis ini tidak bisa terlepas dari bantuan yang diberikan berbagai pihak, baik yang bersifat langsung maupun tidak langsung. Untuk itu, penulis ingin mengucapkan terima kasih yang sebanyak-banyaknya kepada:

1. Prof. Ir. Suyoto, M.Sc., Ph.D., selaku Ketua Program Studi Magister Teknik Informatika serta seluruh dosen dan karyawan yang telah membantu penulis selama menempuh studi.
2. Bapak Ir. A. Djoko Budiyanto, M. Eng., Ph. D. selaku Dosen Pembimbing I, dan bapak Dr. Ir. Alb. Joko Santoso, MT. selaku Dosen Pembimbing II yang telah memberikan banyak sekali masukan, nasihat dan bahkan menjadi teman dalam berdiskusi dalam rangka penyelesaian tulisan ini dan berbagai kebaikan yang diberikan kepada penulis untuk menyelesaikan tesis ini.
3. Bapa dan Mama tercinta atas dukungan doa yang tidak pernah berhenti kepadaku, atas kasih sayang yang berkelimpahan, atas motivasi di saat terasa

lemah, atas perkataan - perkataan yang menguatkan penulis disaat terasa berat dan tidak mampu, maupun bantuan moril dan materil yang penulis terima.

4. Kekasih tercinta di Kupang yang selalu setia mendukung dan memberi motivasi sehingga penulis bisa menyelesaikan tesis ini.
5. Buat teman-teman dan saudara-saudari baik yang di Kupang maupun di Yogyakarta. Terima kasih untuk kebersamaan rasa kekeluargaan dan secara tidak langsung memberikan dukungan semangat buat penulis.
6. Seluruh teman – teman MTF angkatan September 2015 yang telah menjadi sahabat, berbagi suka dan duka selama belajar di Yogyakarta. Kalian semua adalah teman-teman terbaik dalam hidup. Semoga persabatan kita kekal selamanya.

Demikian penulisan tesis ini dibuat dengan sebaik-baiknya oleh penulis. Penulis menyadari bahwa penulisan tesis ini masih memiliki banyak kekurangan dan jauh dari kata sempurna. Oleh karena itu saran dan kritik yang membangun demi penyempurnaan penulisan tesisi ini akan selalu penulis nantikan.

Akhir kata, penulis mengharapkan semoga penulisan tesis ini dapat berguna dan bermanfaat sehingga dapat memberikan inspirasi bagi pembacanya, dan dapat berguna bagi kita semua. Tuhan Yesus Memberkati.

Yogyakarta, Juli 2017

Penulis

DAFTAR ISI

PENGESAHAN TESIS.....	ERROR! BOOKMARK NOT DEFINED.
PENGESAHAN TESIS.....	ERROR! BOOKMARK NOT DEFINED.
PERNYATAAN.....	IV
INTISARI	V
ABSTRACT.....	VI
LEMBAR PERSEMBAHAN	VII
KATA PENGANTAR.....	VIII
DAFTAR ISI.....	X
DAFTAR GAMBAR.....	XII
DAFTAR TABEL	XIII
BAB I PENDAHULUAN.....	ERROR! BOOKMARK NOT DEFINED.
1.1 Latar Belakang	Error! Bookmark not defined.
1.2 Rumusan Masalah	Error! Bookmark not defined.
1.3 Batasan Masalah.....	Error! Bookmark not defined.
1.4 Tujuan Penelitian	Error! Bookmark not defined.
1.5 Manfaat Penelitian	Error! Bookmark not defined.
1.6 Sistematika Penulisan	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA.....	ERROR! BOOKMARK NOT DEFINED.
BAB III LANDASAN TEORI.....	ERROR! BOOKMARK NOT DEFINED.
3.1 Peringkasan Teks Secara Otomatis	Error! Bookmark not defined.
3.2 Stemming	Error! Bookmark not defined.
3.3 Sentence Scoring.....	Error! Bookmark not defined.
3.4 Decision Tree	Error! Bookmark not defined.

3.5	Bahasa Indonesia.....	Error! Bookmark not defined.
3.6	Berita	Error! Bookmark not defined.
BAB IV METODOLOGI PENELITIAN ERROR! BOOKMARK NOT DEFINED.		
4.1	Pengumpulan Dokumen	Error! Bookmark not defined.
4.2	Training Data	Error! Bookmark not defined.
4.2.1	Sentence Scoring.....	Error! Bookmark not defined.
4.2.2	Decision Tree	Error! Bookmark not defined.
4.3	Testing Data	Error! Bookmark not defined.
4.4	Evaluasi	Error! Bookmark not defined.
4.5	Lingkungan Pengembangan	Error! Bookmark not defined.
BAB V HASIL DAN PEMBAHASAN ERROR! BOOKMARK NOT DEFINED.		
5.1	Analisis Data	Error! Bookmark not defined.
5.2	Pohon Keputusan	Error! Bookmark not defined.
5.3	Evaluasi	Error! Bookmark not defined.
BAB VI KESIMPULAN DAN SARAN ERROR! BOOKMARK NOT DEFINED.		
6.1	Kesimpulan	Error! Bookmark not defined.
6.2	Saran.....	Error! Bookmark not defined.
DAFTAR PUSTAKA ERROR! BOOKMARK NOT DEFINED.		

DAFTAR GAMBAR

Gambar 1 Diagram alur peringkasan teks secara otomatis **Error! Bookmark not defined.**

Gambar 2 Konsep Decision Tree **Error! Bookmark not defined.**

Gambar 3 Model pohon keputusan dengan algoritma C4.5 **Error! Bookmark not defined.**

Gambar 4 Tahap Training **Error! Bookmark not defined.**

Gambar 5 Alur pemrosesan tokenisasi..... **Error! Bookmark not defined.**

Gambar 6 Tahap Testing..... **Error! Bookmark not defined.**

Gambar 7 Contoh dokumen teks berita..... **Error! Bookmark not defined.**

Gambar 8 Pohon keputusan **Error! Bookmark not defined.**

Gambar 9 Grafik hasil uji peringkasan teks **Error! Bookmark not defined.**

DAFTAR TABEL

Tabel 1 Contoh data training.....**Error! Bookmark not defined.**

Tabel 2 Hasil pemrosesan awal dokumen teks**Error! Bookmark not defined.**

Tabel 3 Statistik hasil pembobotan kalimat**Error! Bookmark not defined.**

Tabel 4 Pengelompokan bobot.....**Error! Bookmark not defined.**

Tabel 5 Perbandingan ringkasan manual dan ringkasan sistem**Error! Bookmark not defined.**

Tabel 6 Hasil perhitungan precision, recall, dan f-measure**Error! Bookmark not defined.**

Tabel 7 Hasil uji peringkasan teks**Error! Bookmark not defined.**