

**PERINGKASAN TEKS OTOMATIS SECARA EKSTRAKTIF PADA
ARTIKEL BERITA KESEHATAN BERBAHASA INDONESIA
DENGAN MENGGUNAKAN METODE LATENT SEMANTIC
ANALYSIS**

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:

Nurina Savanti Widya Gotami

NIM: 145150207111129



PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2018

PENGESAHAN

PERINGKASAN TEKS OTOMATIS SECARA EKSTRAKTIF PADA ARTIKEL BERITA
KESEHATAN BERBAHASA INDONESIA DENGAN MENGGUNAKAN METODE LATENT
SEMANTIC ANALYSIS

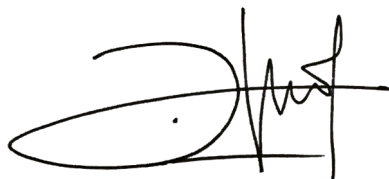
SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :
Nurina Savanti Widya Gotami
NIM: 145150207111129

Skripsi ini telah diuji dan dinyatakan lulus pada
19 Januari 2018
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I



Indriati, S.T., M.Kom
NIP: 19831013 201504 2 002

Dosen Pembimbing II



Ratih Kartika Dewi, S.T., M.Kom
NIK: 201503 890520 2 001

Mengetahui
Ketua Jurusan Teknik Informatika



Tri Astoto Kurniawan, S.T, M.T, Ph.D
NIP: 19710518 200312 1 001

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 1 Januari 2018



Nurina Savanti Widya Gotami

NIM: 145150207111129

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT penulis limpahkan atas rahmat dan karunia-nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul “Peringkasan Teks Otomatis Secara Ekstraktif Pada Artikel Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode *Latent Semantic Analysis*”. Penyusunan skripsi ini merupakan syarat untuk mendapatkan gelar sarjana pada Program Studi Teknik Informatika, Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya.

Di dalam proses penyelesaian skripsi ini penulis banyak mendapatkan bantuan serta dukungan moral maupun material dari berbagai pihak. Sehingga penulis mengucapkan terima kasih sebanyak-banyaknya kepada :

1. Ibu Indriati, S.T, M.Kom., dan Ibu Ratih Kartika Dewi, S.T, M.Kom. selaku dosen pembimbing yang memberikan bimbingan dan mengarahkan penulis dalam penyelesaian skripsi ini.
2. Bapak Wayan Firdaus Mahmudy, S.Si, M.T, Ph.D. selaku Dekan Fakultas Ilmu Komputer Universitas Brawijaya Malang.
3. Bapak Tri Astoto Kurniawan, S.T, M.T, Ph.D., dan Bapak Agus Wahyu Widodo, S.T, M.Cs., selaku Ketua Jurusan Teknik Informatika UB dan Ketua Program Studi Teknik Informatika UB.
4. Orang tua penulis, Bapak Arif Sarjono, S.Pd, M.Si., dan Ibu Puji Yuswanti, Amd.Keb., dan seluruh keluarga atas doa, dukungan moral serta material.
5. Kedua adik saya yang sangat saya sayangi Agesti Veva Kalista dan Septika Nanda Arifia yang selalu mendukung, memberi perhatian, menyemangati dan menghibur serta memberi motivasi dan selalu memberikan keceriaan dihidup saya.
6. Seluruh civitas akademik Fakultas Ilmu Komputer Universitas Brawijaya atas dukungan yang telah diberikan kepada penulis selama menempuh studi di Teknik Informatika Fakultas Ilmu Komputer Universitas Brawijaya.
7. Seluruh dosen Fakultas Ilmu Komputer Universitas Brawijaya yang telah memberikan ilmu yang bermanfaat kepada penulis selama menempuh studi di Teknik Informatika Fakultas Ilmu Komputer Universitas Brawijaya.
8. Sahabat-sahabat saya saudara Hamim Fathul Aziz dan Yane Marita Febrianti, Mahdarani Dwi Laxmi, Yessi Agustin P serta kepada Saudari Kharisna Prayogi Eviana dan Dwi Qunita Putri A.P yang telah banyak membantu dan mendukung penulis dalam penyelesaian skripsi ini.
9. Serta semua pihak yang telah membantu dalam proses penyelesaian skripsi ini yang tak dapat penulis sebutkan satu persatu.

Dalam penulisan skripsi ini, penulis menyadari terdapat banyak kekurangan serta kesalahan, sehingga penulis mengharapkan adanya kritik dan saran yang membangun dari berbagai pihak untuk memperbaiki diri. Penulis juga berharap agar skripsi ini dapat memberikan manfaat.

Malang, 17 Agustus 2017

Nurina Savanti Widya Gotami
nurinaswgotami@gmail.com

ABSTRAK

Meningkatnya jumlah data digital dengan cepat setiap tahunnya, seperti data teks dokumen yang dapat berupa teks artikel berita ini membuat sulit untuk membaca seluruh informasi yang ada. Banyaknya informasi *online* yang tersedia pada penyedia informasi *online* yang mempengaruhi hasil informasi yang diperoleh relatif sulit serta waktu yang lebih lama untuk mengekstraksi suatu informasi pada sebuah dokumen. Oleh karena itu, perlu adanya peringkasan teks otomatis secara ekstraktif pada artikel kesehatan berbahasa Indonesia dengan tujuan membantu pembaca atau pengguna untuk mempermudah dalam proses ekstraksi data atau informasi yang ada pada dokumen dengan waktu yang cepat serta efisien serta dapat menghasilkan informasi penting yang ringkas serta jelas. Pada penelitian ini menggunakan metode *latent semantic analysis* (LSA) merupakan sebuah metode yang mengekstrak struktur semantik atau makna yang tersembunyi pada sebuah kalimat dan menghasilkan ringkasan yang bermakna umum atau luas. Metode LSA menggunakan pendekatan aljabar *linear singular value decomposition* (SVD) dengan membentuk matriks representasi dari asosiasi *term* yang merupakan kata-kata pada dokumen yang berhubungan erat. *Cross LSA* digunakan untuk menyusun urutan ringkasan dalam tahap ekstraksi ringkasan. Pengujian penelitian ini menghasilkan bahwa hasil peringkasan teks dengan metode LSA memperoleh nilai rata-rata akurasi *precision*, *recall* dan *f-measure* secara berurutan pada *compression rate* 50% dengan nilai 0.668, 0.743, 0.700 dan 0.690 sedangkan pada *compression rate* 40% sebesar 0.696, 0.605, 0.642 dan 0.663.

Kata kunci: peringkasan teks ekstraktif, *latent semantic analysis*, *singular value decomposition*.

ABSTRACT

Increasing the amount of digital data quickly each year, such as text data documents that can be the text of this news article makes it difficult to read all the information available. The large number of online information available to online information providers that affect the results of information obtained is relatively difficult as well as a longer time to extract information on a document. Therefore, the need for automatic extracting of extractive texts on health articles in Indonesian language with the aim of helping the reader or user to facilitate the process of extraction of data or information in the document with a fast and efficient time and can generate important information is concise and clear. In this research, the latent semantic analysis (LSA) method is a method that extracts semantic structure or hidden meaning in a sentence and produces a general or broad meaningful summary. The LSA method uses the linear value decomposition (SVD) linear algebra approach by forming a representation matrix of term associations which are words in closely related documents. Cross method LSA is used to compile a summary sequence in the summary extraction stage. Tests of this research resulted that the result of text summary with LSA method obtained the accuracy value of precision, recall and f-measure in consecutive order at compression rate 50% with value 0.668, 0.743, 0.700 and 0.690 while at compression rate 40% equal to 0.696, 0.605, 0.642 and 0.663.

Keyword: Extractive text summarization, latent semantic analysis, singular value decomposition.

DAFTAR ISI

PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN ORISINALITAS	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR KODE PROGRAM	xii
BAB 1 PENDAHULUAN.....	Error! Bookmark not defined.
1.1 Latar belakang.....	Error! Bookmark not defined.
1.2 Rumusan masalah.....	Error! Bookmark not defined.
1.3 Tujuan	Error! Bookmark not defined.
1.4 Manfaat.....	Error! Bookmark not defined.
1.5 Batasan masalah	Error! Bookmark not defined.
1.6 Sistematika pembahasan	Error! Bookmark not defined.
BAB 2 LANDASAN KEPUSTAKAAN	Error! Bookmark not defined.
2.1 Kajian Pustaka	Error! Bookmark not defined.
2.2 Dasar Teori.....	Error! Bookmark not defined.
2.2.1 Peringkasan Teks.....	Error! Bookmark not defined.
2.2.2 <i>Preprocessing</i>	Error! Bookmark not defined.
2.2.3 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	Error! Bookmark not defined.
2.2.4 <i>Latent Semantic Analysis (LSA)</i>	Error! Bookmark not defined.
2.2.5 Evaluasi pengujian hasil ringkasan..	Error! Bookmark not defined.
2.2.6 Parameter <i>Compression Rate</i>	Error! Bookmark not defined.
BAB 3 METODOLOGI	Error! Bookmark not defined.
3.1 Studi Literatur	Error! Bookmark not defined.
3.2 Pengumpulan Data	Error! Bookmark not defined.

3.3 Perancangan	Error! Bookmark not defined.
3.4 Implementasi	Error! Bookmark not defined.
3.5 Pengujian dan Analisis Hasil.....	Error! Bookmark not defined.
3.6 Penarikan Kesimpulan	Error! Bookmark not defined.
BAB 4 Perancangan	Error! Bookmark not defined.
4.1 Perancangan Data	Error! Bookmark not defined.
4.2 Perancangan Algoritme	Error! Bookmark not defined.
4.2.1 <i>Preprocessing</i>	Error! Bookmark not defined.
4.2.2 Perhitungan TF-IDF.....	Error! Bookmark not defined.
4.2.3 Perhitungan LSA	Error! Bookmark not defined.
4.3 Perancangan pengujian	Error! Bookmark not defined.
4.3.1 Pengujian Akurasi.....	Error! Bookmark not defined.
4.3.2 Pengujian Parameter <i>Compression rate</i>	Error! Bookmark not defined.
	defined.
BAB 5 Implementasi	Error! Bookmark not defined.
5.1 Lingkungan Implementasi Sistem	Error! Bookmark not defined.
5.1.1 Lingkungan Perangkat Keras	Error! Bookmark not defined.
5.1.2 Lingkungan Perangkat Lunak	Error! Bookmark not defined.
5.2 Implementasi Data.....	Error! Bookmark not defined.
5.3 Implementasi Algoritme	Error! Bookmark not defined.
5.3.1 Implementasi <i>preprocessing</i>	Error! Bookmark not defined.
5.3.2 Implementasi Perhitungan LSA	Error! Bookmark not defined.
BAB 6 Pengujian dan analisis	Error! Bookmark not defined.
6.1 Pengujian Parameter Compression Rate	Error! Bookmark not defined.
BAB 7 Penutup	Error! Bookmark not defined.
7.1 Kesimpulan.....	Error! Bookmark not defined.
7.2 Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA.....	Error! Bookmark not defined.
LAMPIRAN A DATA INPUT PENGUJIAN	Error! Bookmark not defined.
LAMPIRAN B PERHITUNGAN MANUAL	Error! Bookmark not defined.
LAMPIRAN C EVALUASI PENGUJIAN.....	Error! Bookmark not defined.

DAFTAR TABEL

Tabel 2. 1 Kajian Pustaka.....	Error! Bookmark not defined.
Tabel 2. 2 Tabel <i>Matrix Confusion</i>	Error! Bookmark not defined.
Tabel 4.1 Data Teks Dokumen Artikel Berita	Error! Bookmark not defined.
Tabel 4.2 Data Hasil Proses <i>Parsing</i>	Error! Bookmark not defined.
Tabel 4. 3 Data Hasil Tokenisasi atau <i>Lexing</i>	Error! Bookmark not defined.
Tabel 4. 4 Tabel Hasil <i>Filtering</i>	Error! Bookmark not defined.
Tabel 4. 5 Hasil Proses <i>Stemming</i>	Error! Bookmark not defined.
Tabel 4. 6 Hasil term indeks	Error! Bookmark not defined.
Tabel 4. 7 Hasil Nilai TF	Error! Bookmark not defined.
Tabel 4. 8 Perhitungan Nilai DF.....	Error! Bookmark not defined.
Tabel 4. 9 Hasil nilai IDF	Error! Bookmark not defined.
Tabel 4. 10 Hasil Nilai TF-IDF.....	Error! Bookmark not defined.
Tabel 4. 11 Bobot Kata pada Kalimat	Error! Bookmark not defined.
Tabel 4. 12 Hasil akhir ringkasan.....	Error! Bookmark not defined.
Tabel 4.13 Hasil urutan kalimat berdasar length.....	Error! Bookmark not defined.
Tabel 4. 14 Data Hasil Perhitungan	Error! Bookmark not defined.
Tabel 4. 15 Tabel Kontigensi	Error! Bookmark not defined.
Tabel 4. 16 Rancangan Tabel Uji Coba Parameter Compression Rate	Error! Bookmark not defined.
Tabel 5.1 Lingkungan Perangkat Keras	Error! Bookmark not defined.
Tabel 6. 1 Hasil Pengujian <i>Compression Rate</i>	Error! Bookmark not defined.

DAFTAR GAMBAR

- Gambar 2. 1 Dekomposisi matriks A pada SVD**Error! Bookmark not defined.**
- Gambar 3. 1 Diagram Alir Alur Tahapan Penelitian ..**Error! Bookmark not defined.**
- Gambar 4. 1 Diagram alir tahapan proses algoritme**Error! Bookmark not defined.**
- Gambar 4.2 Diagram Alir Untuk Tahapan *Preprocessing***Error! Bookmark not defined.**
- Gambar 4. 3 Alur Proses Parsing.....**Error! Bookmark not defined.**
- Gambar 4. 4 Tahapan Proses *Lexing* atau Tokenisasi**Error! Bookmark not defined.**
- Gambar 4. 5 Tahapan Proses *Filtering***Error! Bookmark not defined.**
- Gambar 4. 6 Tahapan Proses *Stemming***Error! Bookmark not defined.**
- Gambar 4. 7 Alur Tahapan Proses Perhitungan TF-IDF**Error! Bookmark not defined.**
- Gambar 4. 8 Alur Proses Perhitungan TF**Error! Bookmark not defined.**
- Gambar 4. 9 Alur Proses Perhitungan DF.....**Error! Bookmark not defined.**
- Gambar 4. 10 Alur Tahapan Proses Perhitungan IDF**Error! Bookmark not defined.**
- Gambar 4. 11 Alur Tahapan Proses TF-IDF**Error! Bookmark not defined.**
- Gambar 4. 12 Alur Proses Tahapan Peringkasan Teks**Error! Bookmark not defined.**
- Gambar 4. 13 Alur Tahapan Proses LSA.....**Error! Bookmark not defined.**
- Gambar 6. 1 Nilai Evaluasi Pengujian Parameter Compression Rate 10%**Error! Bookmark not defined.**
- Gambar 6. 2 Nilai Evaluasi Pengujian Parameter Compression Rate 20%**Error! Bookmark not defined.**
- Gambar 6. 3 Nilai Evaluasi Pengujian Parameter Compression Rate 30%**Error! Bookmark not defined.**
- Gambar 6. 4 Nilai Evaluasi Pengujian Parameter Compression Rate 40%**Error! Bookmark not defined.**
- Gambar 6. 5 Nilai Evaluasi Pengujian Parameter Compression Rate 50%**Error! Bookmark not defined.**

DAFTAR KODE PROGRAM

Kode Program 5. 1 <i>Parsing</i>	Error! Bookmark not defined.
Kode Program 5. 2 Tokenisasi dan <i>Filtering</i>	Error! Bookmark not defined.
Kode Program 5. 3 <i>Stemming</i>	Error! Bookmark not defined.
Kode Program 5. 4 Nilai <i>Raw</i> TF dan DF.....	Error! Bookmark not defined.
Kode Program 5. 5 Nilai Log TF dan IDF	Error! Bookmark not defined.
Kode Program 5. 6 Nilai Log TF.IDF	Error! Bookmark not defined.
Kode Program 5. 7 Matriks N	Error! Bookmark not defined.
Kode Program 5. 8 Eigen Value dan Matriks S.....	Error! Bookmark not defined.
Kode Program 5. 9 Matriks <i>S inverse</i> dan Matriks V.	Error! Bookmark not defined.
Kode Program 5. 10 Ekstraksi Ringkasan	Error! Bookmark not defined.

