

**STUDI DAN IMPLEMENTASI ALGORITMA *SUFFIX TREE***  
***CLUSTERING* UNTUK ORGANISASI DATA HASIL**  
***SEARCH ENGINE***

**SKRIPSI**

**Diajukan untuk Memenuhi Sebagian dari**  
**Syarat Memperoleh Gelar Sarjana Komputer**  
**Program Studi Ilmu Komputer**



oleh

**R SENO ANGGORO A**

**0608508**

**PROGRAM STUDI ILMU KOMPUTER**  
**FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS PENDIDIKAN INDONESIA**  
**BANDUNG**  
**2010**

STUDI DAN IMPLEMENTASI ALGORITMA *SUFFIX TREE CLUSTERING*  
UNTUK ORGANISASI DATA HASIL *SEARCH ENGINE*

Oleh:

R Seno Anggoro A

0608508

Disetujui dan Disahkan oleh:

Pembimbing I

Rizky Rahman J., M.Kom

NIP. 197711252006041002

Pembimbing II

Herbert Siregar, MT.

NIP. 197005022008121001

Mengetahui

Ketua Program Studi Ilmu Komputer

FPMIPA UPI

Drs. Heri Sutarno, MT.

NIP. 195607141984031002

## PERNYATAAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dengan ini saya menyatakan bahwa skripsi yang berjudul "Studi dan Implementasi Algoritma *Suffix Tree Clustering* untuk Organisasi Data Hasil *Search Engine*" ini sepenuhnya karya saya sendiri. Tidak Ada bagian di dalamnya yang merupakan plagiat dari karya orang lain dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat. Atas pernyataan ini, saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, 28 Juni 2010

Yang membuat pernyataan,

**R. Seno Anggoro A.**

NIM. 0608508

*Karya sederhana ini kupersembahkan untuk*

*papa Drs. R. Eko Darmanto*

*mama Hartaty Dwiva*

*mbak Rr. Kristi Ardhi Susanti, S.T.*

*mbak Rr. Niken Rahma Titisari*

*abang R. Pandu Dimas Samiaji*



*"Success ... seems to be connected with action. Successful men keep moving. They make mistakes, but they don't quit."*

-- Conrad Hilton

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya kepada penulis sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul "Studi dan Implementasi Algoritma *Suffix Tree Clustering* untuk Organisasi Data Hasil *Search Engine*" untuk memenuhi sebagian syarat memperoleh gelar Sarjana Komputer pada Program Studi Ilmu Komputer Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia.

Penulis mendapatkan dukungan dan bantuan dari banyak pihak dalam menyelesaikan karya ilmiah ini. Untuk itu penulis menghaturkan banyak terima kasih kepada:

1. Bapak Rizky Rahman J., M.Kom., selaku Pembimbing I sekaligus Pembimbing Akademik, yang telah memberikan arahan, bimbingan, saran dan koreksi baik selama studi maupun selama masa penulisan skripsi.
2. Bapak Herbert Siregar, M.T., selaku Pembimbing II, yang telah bersedia meluangkan waktu dan bersedia memberikan arahan, bimbingan, saran dan koreksi selama masa penulisan skripsi.
3. Bapak Asep Wahyudin, M.T., Jajang Kusnendar, M.T., dan Wahyudin, M.T., yang telah memberikan arahan, bimbingan, saran, dan koreksi selama penulisan skripsi.

4. Seluruh dosen dan staf FPMIPA UPI terutama Program Studi Ilmu Komputer yang telah memberikan ilmu pengetahuan kepada penulis selama masa studi.
5. Kedua orang tua (Drs. R. Eko Darmanto dan Hartaty Dwiva), kakak (Kristi Ardhi Susanti S.T., Niken Rahma Titisari, dan Pandu Dimas Samiaji), dan keluarga besar (alm) Mansyur PS. serta (alm) R. Soekadi, yang selalu memberikan kasih sayang, doa, bimbingan, perhatian, motivasi, dan semangat kepada penulis.

Semoga segala amal kebaikan ini diterima dan dibalas Allah SWT dengan balasan yang baik. Amin.

Tak lupa penulis berterima kasih kepada rekan-rekan seperjuangan dan sahabat-sahabat (Andri Suryadi, Asep Saiful Milah, Dian Nurdiana, Mubarak, dan Agus Sakti Sunarto) yang telah banyak membantu penulis selama studi dan penulisan skripsi, kepada Arnie Rieska Amanda yang penulis banggakan dengan segala kelebihan dan kekurangannya, terima kasih atas dukungan dan semangat yang diberikan selama ini, dan seluruh pihak yang tidak dapat disebutkan namanya satu persatu, yang secara langsung ataupun tidak langsung banyak membantu dalam penyusunan skripsi ini. Semoga apa yang telah kita lakukan bersama dicatat sebagai amal kebaikan oleh Allah SWT. Amin.

Penulis menyadari bahwa apa yang dipaparkan dalam skripsi ini masih jauh dari kesempurnaan. Oleh karena itu penulis akan sangat terbuka menerima saran dan kritik yang membangun dari para pembaca.

Semoga segala hal yang dipaparkan dalam karya ilmiah ini dapat memberikan manfaat bagi kita semua. Amin.

Bandung, Juni 2010

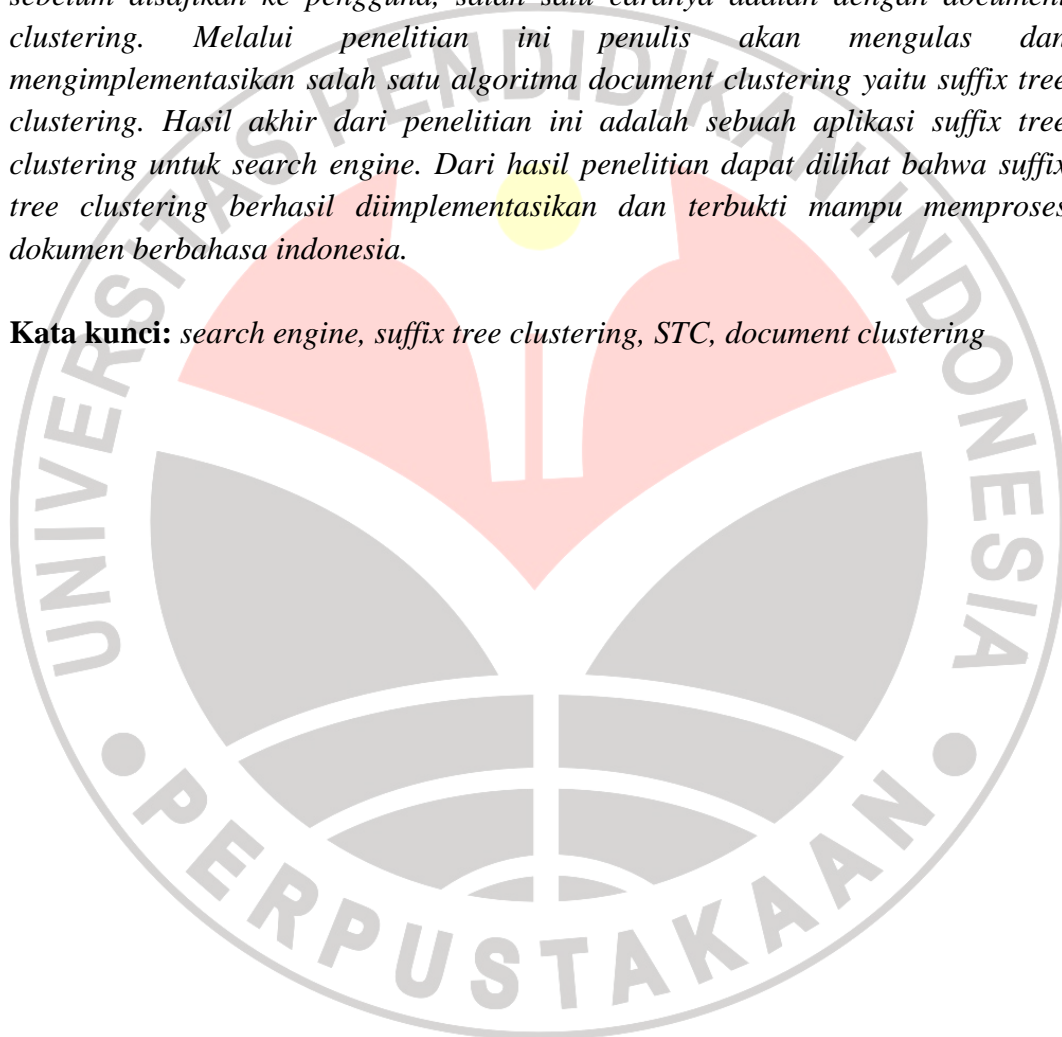
Penulis



## ABSTRAK

*Dengan semakin meningkatnya jumlah data yang ada di internet, search engine atau mesin pencari akan semakin memegang peran pentingnya sebagai alat bantu dalam hal eksplorasi informasi. Namun search engine yang ada saat ini cenderung belum bisa memberikan kemudahan dari sisi eksplorasi informasi. Oleh karena itu perlu dilakukan pengorganisasian data-data hasil search engine sebelum disajikan ke pengguna, salah satu caranya adalah dengan document clustering. Melalui penelitian ini penulis akan mengulas dan mengimplementasikan salah satu algoritma document clustering yaitu suffix tree clustering. Hasil akhir dari penelitian ini adalah sebuah aplikasi suffix tree clustering untuk search engine. Dari hasil penelitian dapat dilihat bahwa suffix tree clustering berhasil diimplementasikan dan terbukti mampu memproses dokumen berbahasa indonesia.*

**Kata kunci:** *search engine, suffix tree clustering, STC, document clustering*



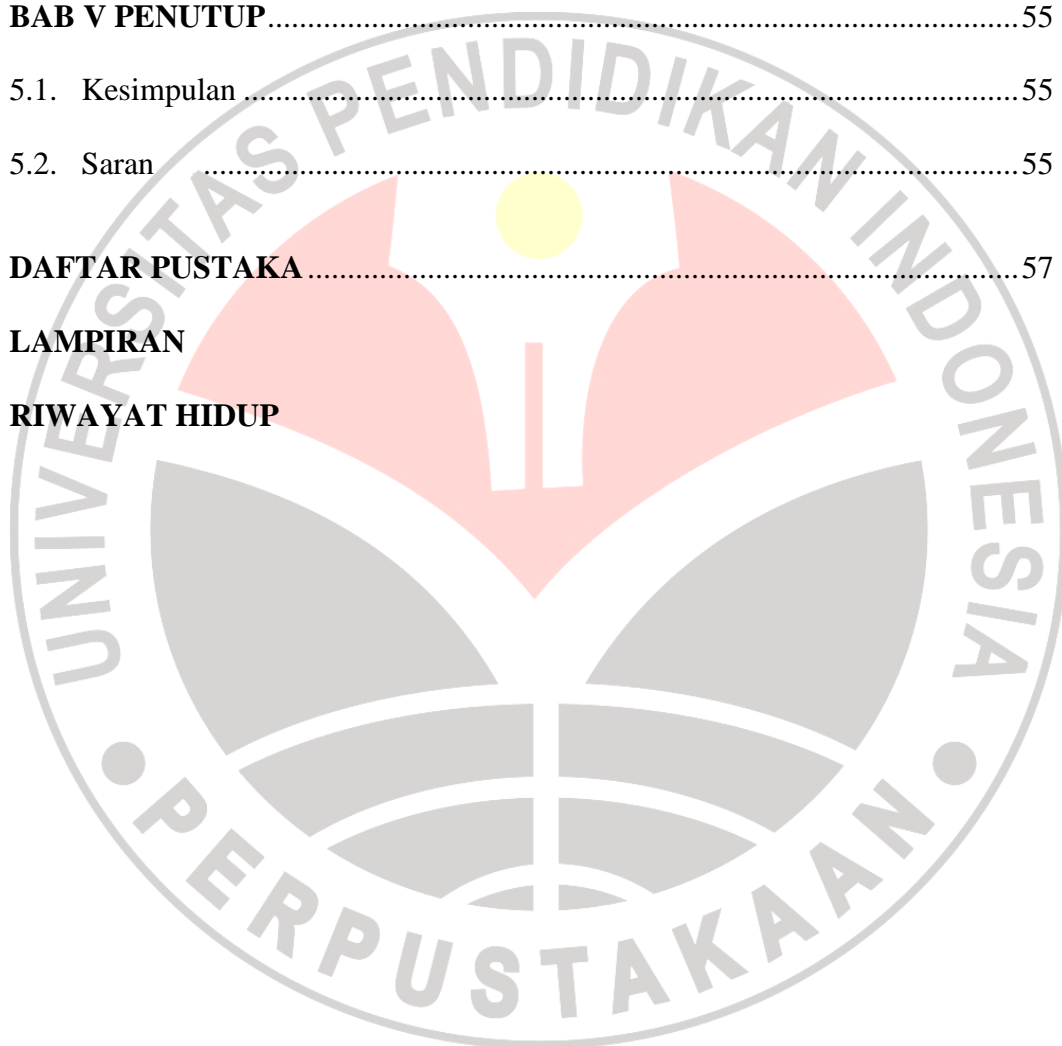


## DAFTAR ISI

<b>KATA PENGANTAR</b> .....	iv
<b>ABSTRAK</b> .....	vii
<b>DAFTAR ISI</b> .....	viii
<b>DAFTAR TABEL</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xii
<b>DAFTAR ISTILAH</b> .....	xiv
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	6
1.3 Batasan Masalah.....	6
1.4 Tujuan .....	7
1.5 Manfaat .....	7
1.6 Sistematika Penulisan .....	8
<b>BAB II TINJAUAN PUSTAKA</b> .....	9
2.1 Data Mining .....	9
2.1.1 Tahap-Tahap pada Data Mining .....	10
2.1.2 Text Mining.....	12
2.2 Metode Clustering.....	12
2.3 Document Clustering .....	14
2.4 Preprocessing .....	16

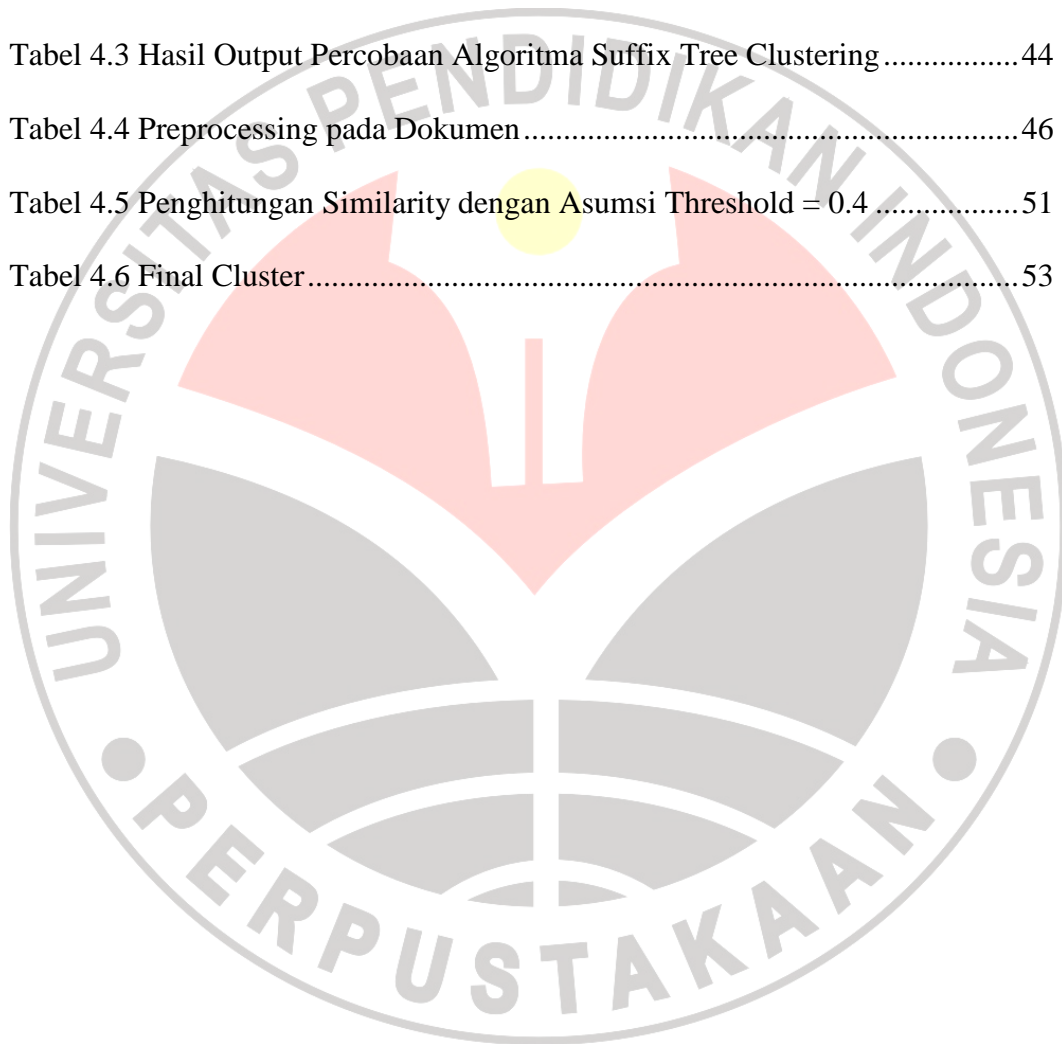
2.5 Suffix Tree Clustering.....	16
2.5.1 Document Cleaning.....	18
2.5.2 Identifikasi Base Cluster.....	19
2.5.3 Mengkombinasikan Base Cluster.....	24
<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>27</b>
3.1 Pendekatan Penelitian.....	27
3.2 Alat dan Bahan Penelitian.....	28
3.2.1 Bahan Penelitian.....	28
3.2.2 Alat Penelitian.....	29
3.3 Desain Penelitian.....	29
3.4 Metode Pengumpulan Data.....	31
3.5 Metode Pengembangan Perangkat Lunak.....	31
3.6 Implementasi.....	33
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....</b>	<b>34</b>
4.1 Studi Kasus.....	34
4.2 Pembangunan Perangkat Lunak.....	34
4.2.1 Batasan Perangkat Lunak.....	34
4.2.2 Analisis Kebutuhan Perangkat Lunak.....	35
4.2.3 Pemodelan Kebutuhan.....	36
4.2.4 Desain Perangkat Lunak.....	38
4.2.5 Implementasi (Coding).....	39
4.2.6 Pengujian.....	40

4.3 Hasil Document Clustering .....	41
4.4 Pembahasan.....	44
4.4.1 Pembentukan Suffix Tree .....	46
4.4.2 Penggabungan Base Cluster .....	51
<b>BAB V PENUTUP.....</b>	<b>55</b>
5.1. Kesimpulan .....	55
5.2. Saran .....	55
<b>DAFTAR PUSTAKA .....</b>	<b>57</b>
<b>LAMPIRAN</b>	
<b>RIWAYAT HIDUP</b>	



## DAFTAR TABEL

Tabel 2.1 Base Cluster dari Contoh pada Gambar 2.6.....	23
Tabel 4.1 Pelaksanaan Pengujian.....	41
Tabel 4.2 Parameter Input.....	42
Tabel 4.3 Hasil Output Percobaan Algoritma Suffix Tree Clustering.....	44
Tabel 4.4 Preprocessing pada Dokumen.....	46
Tabel 4.5 Penghitungan Similarity dengan Asumsi Threshold = 0.4.....	51
Tabel 4.6 Final Cluster.....	53



## DAFTAR GAMBAR

Gambar 1.1 Document Retrieval System Menggunakan Google .....	2
Gambar 1.2 Grafik Perbandingan Algoritma Suffix Tree Clustering .....	5
Gambar 2.1 Proses Data Mining .....	11
Gambar 2.2 Contoh Clustering .....	13
Gambar 2.3 Taksonomi Metode Clustering di Dalam Data Mining .....	14
Gambar 2.4 Contoh Hasil Pencarian Menggunakan Algoritma STC .....	18
Gambar 2.5 Document Cleaning Dan Preprocessing .....	19
Gambar 2.6 Contoh Suffix Tree .....	20
Gambar 2.7 Grafik Nilai Fungsi .....	24
Gambar 2.8 Phrase Cluster Graph .....	26
Gambar 3.1 Siklus Hidup Penelitian .....	27
Gambar 3.2 Model Sekuensial Linier .....	32
Gambar 4.1 Arsitektur Sistem Document Clustering Untuk Search Engine .....	36
Gambar 4.2 Diagram Use Case Perangkat Lunak Suffix Tree Clustering Untuk Search Engine .....	37
Gambar 4.3 Basis Data Sistem Search Engine Yang Akan Diolah .....	38
Gambar 4.4 Hasil Percobaan Dengan 50 Dokumen Yang Dipilih Secara Acak .....	43
Gambar 4.5 Implementasi Algoritma Suffix Tree Clustering .....	45
Gambar 4.6 Struktur Suffix Tree Setelah Dokumen 1 Diproses .....	47
Gambar 4.7 Kondisi Suffix Tree Setelah Dokumen 2 Diproses .....	48

Gambar 4.8 Struktur Suffix Tree Setelah Seluruh Dokumen Diproses .....	49
Gambar 4.9 Base Cluster yang Terbentuk Beserta Skornya .....	50
Gambar 4.10 Contoh Phrase Cluster Graph Dimana Terbentuk 2 Kelompok Cluster.....	53



## DAFTAR ISTILAH

### ***Black box testing***

Teknik pengujian perangkat lunak dimana pengujian dilakukan pada bagaimana sistem menerima input dan mengeluarkan output yang diinginkan, tidak memperhitungkan apa yang terjadi di dalam sistem.

### ***Class diagram***

Diagram yang mewakili kelas, bagian dari kelas dan menggambarkan hubungan satu kelas dengan kelas yang lain.

### ***OOP***

*Object oriented programming*, adalah paradigma pemrograman yang menggunakan objek untuk mendesain aplikasi atau program komputer.

### ***Query***

Kumpulan data yang dikembalikan berdasarkan *term*.

### ***Resource***

Entitas fisik atau virtual yang terpakai dalam menjalankan suatu aplikasi komputer.

### ***Similarity***

Tingkat kemiripan antar dua objek.

## **STC**

*Suffix tree clustering, algoritma document clustering.*

## **String**

Untaian kata sebagai representasi dari dokumen atau kalimat.

## **Term**

Kata kunci yang digunakan dalam proses pencarian pada *search engine*.

## **Use case diagram**

Diagram yang menunjukkan fungsionalitas suatu sistem atau kelas dan bagaimana sistem tersebut berinteraksi dengan dunia luar dan menjelaskan sistem secara fungsional yang terlihat user.

## **Web Crawler**

Program komputer yang secara otomatis berfungsi menyalin konten halaman web yang dikunjungi.