

**TEMU KEMBALI INFORMASI PADA PENCARIAN  
JURNAL SKRIPSI MENGGUNAKAN METODE SINGLE  
PASS CLUSTERING**

**SKRIPSI**



**OLEH:**

**RIMA NOER SAFITRI**

**09.621.066**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH GRESIK  
2013**

**TEMU KEMBALI INFORMASI PADA PENCARIAN JURNAL  
SKRIPSI MENGGUNAKAN METODE SINGLE PASS  
CLUSTERING**

**SKRIPSI**

Diajukan sebagai syarat memperoleh gelar Sarjana Komputer  
Program Studi Teknik Informatika jenjang S-1 Fakultas Teknik  
Universitas Muhammadiyah Gresik



**OLEH:**

**RIMA NOER SAFITRI**

**09.621.066**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH GRESIK  
2013**

## **PERNYATAAN KEASLIAN SKRIPSI**

Yang bertandatangan dibawah ini:

Nama : Rima Noer Safitri

NIM : 09 621 066

Fakultas/Program studi : Teknik/Teknik Informatika

Menyatakan dengan sebenarnya bahwa skripsi yang berjudul “**TEMU KEMBALI INFORMASI PADA PENCARIAN JURNAL SKRIPSI MENGGUNAKAN METODE SINGLE PASS CLUSTERING**” yang saya buat untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Gresik, bukan merupakan tiruan atau duplikasi dari skripsi yang sudah dipublikasikan dan atau pernah dipakai untuk mendapatkan gelar kesarjanaan di lingkungan Universitas Muhammadiyah Gresik maupun di Perguruan Tinggi atau Instansi manapun, kecuali bagian yang secara tertulis dikutip dalam naskah ini atau disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di kemudian hari, saya terbukti melanggar pernyataan saya tersebut di atas, maka saya bersedia untuk mempertanggungjawabkan serta diproses sesuai peraturan yang berlaku.

Gresik, 14 Februari 2013

RIMA NOER SAFITRI

NIM 09 621 066

**LEMBAR PERSETUJUAN**

**SIDANG SKRIPSI**

**TEMU KEMBALI INFORMASI PADA PENCARIAN JURNAL SKRIPSI  
MENGUNAKAN METODE SINGLE PASS CLUSTERING**

Oleh

**RIMA NOER SAFITRI**

**NIM 09.621.066**

Disetujui untuk dipresentasikan dalam sidang skripsi.

**Susunan Tim Pembimbing**

Pembimbing I

Pembimbing II

Harunur Rosyid, S.T., M.Kom.

Eko Prasetyo, S.Kom., M.Kom.

NIP 06 210 408 106

NIP 06 240 511 015

Diterima pada tanggal ..... dan dinyatakan telah memenuhi seluruh persyaratan sidang skripsi.

Mengetahui,

Ketua Program Studi Teknik Informatika

Universitas Muhammadiyah Gresik

Soffiana Agustin, S.Kom., M.Kom.

NIP: 197711292005012001

# LEMBAR PENGESAHAN

## SKRIPSI

### TEMU KEMBALI INFORMASI PADA PENCARIAN JURNAL SKRIPSI MENGUNAKAN METODE SINGLE PASS CLUSTERING

Oleh

**RIMA NOER SAFITRI**

**NIM 09.621.066**

Telah dipertahankan di depan Tim Penguji

Pada tanggal 18 Februari 2013

#### Susunan Tim Penguji

Penguji I (Ketua)

Penguji II (Sekretaris)

Utomo Pujiyanto, S.Kom., M.Kom

Harunur Rosyid, S.T., M.Kom

NIP. 06 240 508 010

NIP. 06 210 408 106

Penguji III (Anggota)

Penguji IV (Anggota)

Soffiana Agustin, S.Kom., M.Kom.

Eko Prasetyo, S.Kom., M.Kom.

NIP. 197711292005012001

NIP. 06 240 511 015

Diterima dan dinyatakan memenuhi syarat kelulusan pada tanggal.....

Mengetahui,

Dekan Fakultas Teknik  
Universitas Muhammadiyah Gresik

Ketua Program Studi Teknik Informatika  
Universitas Muhammadiyah Gresik

Moch. Nuruddin, S.T., M.T.

NIP: UMG 06 119 810 043

Soffiana Agustin, S. Kom., M. Kom.

NIP: 197711292005012001

## KATA PENGANTAR

Dengan menyebut asma Allah Yang Maha Pengasih dan Maha Penyayang, puji syukur kehadirat Allah SWT yang hanya atas ijin-Nya penulis dapat menyelesaikan skripsi dengan judul *“Temu Kembali Informasi pada Pencarian Jurnal Skripsi menggunakan Metode Single Pass Clustering”*.

Terselesaikannya skripsi ini dengan baik berkat dukungan, motivasi, petunjuk, bimbingan dan do'a dari berbagai pihak. Oleh karena itu, iringan do'a dan ucapan terimakasih yang sebesar-besarnya penulis sampaikan kepada:

1. Ibuku, Nurul Chamimah yang selalu mendo'akan, mendidik, membina, memberikan segala pengorbanannya yang tak terhitung, serta menaruh harapan kepada penulis, serta atas restu dan do'a dari beliau sehingga penulis bisa menyelesaikan skripsi ini dan lulus 3.5 tahun sesuai dengan yang penulis harapkan.
2. Ayahku, Ribuang yang telah mendidik, berkorban dan atas kerja kerasnya sehingga penulis bisa sampai menjadi seorang sarjana seperti saat ini.
3. Bapak Utomo Pujianto,S.Kom,.M.Kom selaku dosen penguji I, yang telah memberikan masukan dan arahan bagi penulis dalam penyelesaian skripsi, serta telah memberikan ilmu dan pengalaman bagi penulis selama dibangku perkuliahan.
4. Bpk. Harunur Rasyid,S.Kom,.M.Kom selaku pembimbing I dan Bpk. Eko Prasetyo,S.Kom,.M.Kom selaku Pembimbing II, yang senantiasa meluangkan waktunya untuk memberikan bimbingan, arahan serta masukan bagi penulis dalam penyelesaian skripsi.
5. Bu Soffiana Agustin,S.Kom,.M.Kom selaku dosen penguji II, yang telah memberikan masukan serta arahan bagi penulis dalam penyelesaian skripsi.
6. Pak Deni Sutaji,S.Kom, serta semua dosen di Program Studi Teknik Informatika yang telah memberikan ilmu dan pengetahuan kepada

penulis selama di bangku perkuliahan, serta dorongan sehingga penulis dapat segera menyelesaikan skripsi ini.

7. Teman-teman seperjuangan di Program Studi Teknik Informatika UMG angkatan 2009.
8. Sahabat dan Keluarga besar Departemen Keputrian LDK UKKI UMG yang telah memberi dukungan dan do'a.

Penulis menyadari dalam penyusunan skripsi ini masih banyak kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun dari pembaca. Semoga skripsi ini bermanfaat tidak hanya bagi penulis khususnya, tetapi juga bermanfaat untuk pembaca pada umumnya.

Gresik, Februari 2013

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN SAMPUL LUAR</b> .....	ii
<b>HALAMAN SAMPUL DALAM</b> .....	iii
<b>LEMBAR PERNYATAAN</b> .....	iv
<b>LEMBAR PERSETUJUAN</b> .....	v
<b>LEMBAR PENGESAHAN</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	xi
<b>DAFTAR TABEL</b> .....	xiii
<b>DAFTAR PERSAMAAN</b> .....	xiv
<b>DAFTAR KODE PROGRAM</b> .....	xv
<b>DAFTAR LAMPIRAN</b> .....	xvi
<b>INTISARI</b> .....	xvii
<b>ABSTRACT</b> .....	xviii
<b>BAB I PENDAHULUAN</b>	
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Tujuan Penelitian .....	3
1.4. Batasan Masalah .....	3
1.5. Metodologi Penelitian .....	3
1.6. Sistematika Penulisan .....	4
1.7. Penjadwalan Kegiatan Penelitian .....	5
<b>BAB II LANDASAN TEORI</b>	
2.1. Tinjauan Studi .....	6
2.2. Tinjauan Pustaka .....	7
2.2.1. Jurnal Skripsi .....	7
2.2.2. Sistem Temu Kembali Informasi .....	8
2.2.3. Single Pass Clustering .....	21



2.2.4. Evaluasi .....	24
<b>BAB III ANALISIS DAN PERANCANGAN SISTEM</b>	
3.1. Analisis Sistem.....	27
3.1.1. Analisis Masalah .....	27
3.1.2. Analisis Spesifikasi Kebutuhan Sistem .....	28
3.1.3. Analisis Kebutuhan Fungsional Sistem .....	28
3.1.4. Deskripsi Sistem .....	29
3.2. Perancangan Sistem .....	30
3.2.1. Flowchart Sistem.....	30
3.2.2. Context Diagram Sistem .....	42
3.2.3. Diagram Berjenjang .....	43
3.2.4. Data Flow Diagram .....	45
3.2.5. Contoh Perhitungan Manual .....	49
3.2.6. Perancangan Database .....	60
3.2.7. Perancangan Interface .....	64
3.2.8. Perancangan Uji Coba.....	68
<b>BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM</b>	
4.1. Implementasi .....	69
4.1.1. Implementasi <i>Indexing</i> .....	69
4.1.2. Implementasi <i>Formulasi Query</i> .....	76
4.1.3. Implementasi Pencarian .....	76
4.1.4. Implementasi Single Pass Clustering .....	79
4.2. Implementasi Antarmuka .....	81
4.3. Pengujian Sistem .....	84
4.4. Pembahasan .....	93
<b>BAB V PENUTUP</b>	
5.1. Kesimpulan .....	95
5.2. Saran.....	96
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 2.1 Proses Sistem Temu Kembali Informasi .....	11
Gambar 2.2 Kerangka dari Sistem Temu Kembali Informasi .....	12
Gambar 2.3 Algoritma Single Pass Clustering .....	22
Gambar 2.4 Diagram Alir Algoritma Single Pass Clustering .....	23
Gambar 2.5 Single Pass Clustering .....	24
Gambar 2.6 Recall dan Precision pada contoh hasil temu kembali informasi ....	25
Gambar 3.1 Flowchart Utama Sistem .....	31
Gambar 3.2 Flowchart proses indexing abstrak skripsi .....	32
Gambar 3.3 Flowchart formulasi query .....	33
Gambar 3.4 Flowchart pencarian .....	35
Gambar 3.5 Diagram Alir Proses Preprocessing .....	36
Gambar 3.6 Diagram Alir case folding dan tokenizing .....	37
Gambar 3.7 Diagram Alir Proses Filtering .....	38
Gambar 3.8 Diagram Alir Algoritma Stemming Nazief-Andriani .....	39
Gambar 3.9 Diagram Alir Proses Perhitungan Bobot dan simmilarity antar dokumen dan query .....	40
Gambar 3.10 Diagram Alir Proses Algoritma Single Pass Clustering .....	41
Gambar 3.11 Context Diagram Sistem Pencarian Jurnal Skripsi .....	43
Gambar 3.12 Diagram Berjenjang Sistem Pencarian Jurnal Skripsi .....	44
Gambar 3.13 Data Flow Diagram Level 1 .....	45
Gambar 3.14 DFD Level 2 Proses Indexing .....	47
Gambar 3.15 DFD Level 2 Proses Formulasi Query .....	47
Gambar 3.16 DFD Level 2 Proses Pencarian .....	48
Gambar 3.17 ERD Sistem Pencarian Jurnal Skripsi .....	60
Gambar 3.18 Antarmuka Halaman Awal Sistem .....	64
Gambar 3.19 Antarmuka Halaman Sistem untuk Lihat seluruh Abstrak yang ada .....	65
Gambar 3.20 Antarmuka Halaman Pencarian .....	66
Gambar 3.21 Antarmuka Halaman Hasil Pencarian .....	67

Gambar 4.1 Antarmuka Halaman Awal Aplikasi Pencarian Jurnal Skripsi .....	81
Gambar 4.2 Antarmuka Halaman Show Corpus.....	82
Gambar 4.3 Antarmuka Halaman Pencarian.....	82
Gambar 4.4 Antarmuka Halaman Hasil Pencarian .....	83
Gambar 4.5 Tampilan Sistem jika Nilai Threshold yang dimasukkan bukan 0-0.2 .....	84
Gambar 4.6 Grafik Hubungan Threshold dengan nilai Precision, Recall dan Accuracy .....	94

## DAFTAR TABEL

Tabel 1.1 Tabel Jadwal Penelitian .....	5
Tabel 2.1 Tabel Kombinasi Awalan Akhiran yang tidak diijinkan .....	16
Tabel 2.2 Tabel Cara Menentukan Tipe Awalan untuk awalan “te-”.....	16
Tabel 2.3 Tabel Jenis Awalan Berdasarkan Tipe Awalannya.....	17
Tabel 2.4 Tabel Ketergantungan .....	25
Tabel 3.1 Tabel Perhitungan TF-IDF.....	55
Tabel 3.2 Tabel Perhitungan Pembobotan .....	56
Tabel 3.3 Tabel Perhitungan dot-product document dengan query .....	57
Tabel 3.4 Tabel Perhitungan vector document-query dan cosine .....	58
Tabel 3.5 Tabel Hasil Perankingan Dokumen .....	59
Tabel 3.6 Struktur Tabel tabel_skripsi .....	61
Tabel 3.7 Struktur Tabel tabel_stopword .....	61
Tabel 3.8 Struktur Tabel tabel_kata_dasar .....	62
Tabel 3.9 Struktur Tabel tabel_index .....	62
Tabel 3.10 Struktur Tabel tabel_query.....	63
Tabel 3.11 Struktur Tabel tabel_hitung_cosine .....	63
Tabel 4.1 Contoh Hasil Implementasi Formulasi Query .....	76
Tabel 4.2 Contoh Hasil Implementasi Proses Perhitungan Cosine Similaritas .....	78
Tabel 4.3 Contoh Hasil Implementasi Perankingan .....	79
Tabel 4.4 Matriks Konfusi Penilaian Relevan atau Tidak Relevannya Keluaran Sistem .....	85
Tabel 4.5 Hasil Pengujian untuk Threshold = 0 .....	90
Tabel 4.6 Hasil Pengujian untuk Threshold = 0.05 .....	90
Tabel 4.7 Hasil Pengujian untuk Threshold = 0.1 .....	91
Tabel 4.8 Hasil Pengujian untuk Threshold = 0.15 .....	91
Tabel 4.9 Hasil Pengujian untuk Threshold = 0.2 .....	92
Tabel 4.10 Hasil Pengujian untuk Threshold = 0.25 .....	92
Tabel 4.11 Nilai Rerata Precision, Recall dan Akurasi pada tiap nilai Threshold .....	93

## DAFTAR PERSAMAAN

Persamaan 2.1 Logaritmic TF .....	19
Persamaan 2.2 Augmented TF .....	20
Persamaan 2.3 IDF .....	20
Persamaan 2.4 Pembobotan .....	20
Persamaan 2.5 Simmilarity Cosine .....	21
Persamaan 2.6 Precission .....	25
Persamaan 2.7 Recall .....	25
Persamaan 2.8 Accuracy .....	25
Persamaan 3.1 Perkalian Vektor Dokumen dengan Query .....	58

## DAFTAR KODE PROGRAM

Kode Program 4.1 Proses Case-folding .....	69
Kode Program 4.2 Proses Tokenizing .....	70
Kode Program 4.3 Proses Stopword-removal .....	70
Kode Program 4.4 Proses Filtering .....	71
Kode Program 4.5 Proses Stemming Cek Kata Dasar .....	71
Kode Program 4.6 Proses Stemming Penghapusan Akhiran Infleksi .....	72
Kode Program 4.7 Proses Stemming Penghapusan Akhiran Derrivasi -i, -an, -kan .....	73
Kode Program 4.8 Proses Indexing .....	74
Kode Program 4.9 Proses Pembobotan .....	75
Kode Program 4.10 Kode Program Perhitungan Panjang Vector Dokumen Abstrak .....	77
Kode Program 4.11 Kode Program Perkalian Vector Dokumen Abstrak dengan Query.....	77
Kode Program 4.12 Kode Program Hitung Cosine Simmilarity.....	78
Kode Program 4.13 Kode Program Proses Pengambilan dan Perepresentasian abstrak ranking satu .....	79
Kode Program 4.14 Kode Program Proses Perhitungan Simmilaritas dalam Single Pass Clustering .....	80

## **DAFTAR LAMPIRAN**

- Lampiran 1 : Lampiran Daftar Stopword
- Lampiran 2 : Lampiran Data Abstrak Skripsi yang digunakan dalam Pengujian
- Lampiran 3 : Lampiran Hasil Implementasi Proses Case-folding
- Lampiran 4 : Lampiran Hasil Implementasi Proses Tokenizing
- Lampiran 5 : Lampiran Hasil Implementasi Proses Filtering dan stemming
- Lampiran 6 : Lampiran Source code untuk proses preprocessing dokumen
- Lampiran 7 : Lampiran Source Code untuk Proses Stemming
- Lampiran 8 : Lampiran Source Code Proses Indexing
- Lampiran 9 : Lampiran Source code Proses Formulasi Query
- Lampiran 10 : Lampiran Source code Proses Hhitung Kemiripan
- Lampiran 11 : Lampiran Source code Metode Single Pass Clustering
- Lampiran 12 : Lampiran Source code Program Utama

**INFORMATION RETRIEVAL THE JURNAL THESIS USING SINGLE  
PASS CLUSTERING**

By

**RIMA NOER SAFITRI**

**09 621 066**

Information submitted to the Faculty of Engineering Program  
Muhammadiyah University of Gresik on February 2013 to meet most  
requirements of obtaining an undergraduate degree S-1 Engineering Program  
Information

**ABSTRACT**

Journal thesis is often used as a source of reference in making a study. The greater the number of thesis collection owned by the speed and accuracy of information is very important for the acquisition of information seekers.

This study seeks to apply information retrieval with single pass clustering methods for finding results in the form of an abstract thesis that has relevance to the user query input.

System testing conducted empirical tests with 13 queries and threshold input from 0-0.25. The data used for the test system is a collection of abstracts of these documents Informatics Engineering Program UMG years 2007-2012 were taken 73 documents from digital library UMG abstract document. The selection of the threshold value will determine the performance of the system. Based on the test results obtained are quite ideal threshold value is 0:15 with precision value 74.4505%, recall value reached 68.0769% and 96.1012 % accuracy achieved.

**Keywords:** *Information Retrieval, Single Pass Clustering, Journal Abstract Thesis.*

Supervisor : Harunur Rosyid, S.Kom.,M.Kom.

Co. Supervisor : Eko Prasetyo,S.Kom.,M.Kom.