

**RANCANG BANGUN APLIKASI DOKUMEN *CLUSTERING*  
DENGAN METODE K-MEANS BERDASARKAN *WINNOWERING*  
*FINGERPRINT SIMILARITY***

**TUGAS AKHIR**

Diajukan Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Teknik  
Pada Jurusan Teknik Informatika

Oleh

**RIO WILYANTO**  
**10751000020**



UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU  
PEKANBARU  
2014**

## LEMBAR PENGESAHAN

### RANCANG BANGUN APLIKASI DOKUMEN CLUSTERING DENGAN METODE K-MEANS BERDASARKAN WINNOWER FINGERPRINT SIMILARITY

#### TUGAS AKHIR

Oleh

**RIO WILYANTO**  
**10751000020**

Telah dipertahankan di depan sidang dewan penguji  
Sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika  
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau  
Di Pekanbaru, pada tanggal 24 Juni 2014

Pekanbaru, 26 Juni 2014

Mengesahkan,



Dekan

**Dra. Hj. Yenita Morena, M.Si**  
**NIP. 19601125 198503 2 002**

Ketua Jurusan

**Elin Haerani, ST., M.Kom**  
**NIP. 19810523 200710 2 003**

#### DEWAN PENGUJI

Ketua : Drs. Martius, M.Hum  
Sekretaris : Surya Agustian, ST., M.Kom  
Penguji I : Jasril, S.Si., M.Sc  
Penguji II : Elin Haerani, ST., M.Kom

# RANCANG BANGUN APLIKASI DOKUMEN *CLUSTERING* DENGAN METODE K-MEANS BERDASARKAN *WINNOWER FINGERPRINT SIMILARITY*

**RIO WILYANTO**

**10751000020**

Tanggal Sidang: 24 Juni 2014

Periode Wisuda: November 2014

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

Jalan Soebrantas No. 155 Pekanbaru

## **ABSTRAK**

*Text Mining* disebut juga *text data mining (TDM)* atau *Knowledge Discovery in Text* adalah penemuan informasi baru yang sebelumnya tidak diketahui oleh komputer, dengan cara otomatis mengekstraksi informasi dari sumber-sumber teks yang berbeda. Pada sistem *Text Mining* dikenal berbagai metode dalam pengelompokan suatu teks, salah satunya *Clustering*. *K-means*. *K-means* adalah salah satu metode pengklusteran dengan pendekatan partisi yang mempartisi data yang ada ke dalam bentuk satu atau lebih kelompok. Untuk metode pengujian *clustering* digunakan algoritma *winning*, algoritma ini berfungsi untuk membentuk dokumen teks menjadi kumpulan nilai *fingerprint*. Dalam metode *k-means* digunakan *Jaccard Coefficient* dan *Euclidian Distance* untuk mengukur tingkat *Similarity* antar dokumen. Dari hasil pengujian *relevance jaccard coefficient* didapatkan untuk 4 kelompok sebanyak 73,5 %, pada 7 kelompok 79 %, dan untuk 5 kelompok 96,66%, sementara *relevance euclidian distance* didapatkan untuk 4 kelompok 49,5%, pada 7 kelompok *relevance* gagal, dan 5 kelompok 58,33%. Hasil akhir pengujian ini didapatkan metode *similarity* yang lebih baik dengan menggunakan *jaccard coefficient*.

**Kata Kunci:** Clustering, Euclidian Distance, Jaccard Coefficient, K-means, Relevance, Similarity, Text Mining.

# **APPLICATION DESIGN OF DOCUMENT CLUSTERING WITH K-MEANS METHOD BASED FINGERPRINT WINNOWING SIMILARITY**

**RIO WILİYANTO**

**10751000020**

Final Exam Date: 24 June 2014

Graduation Ceremony Period: November 2014

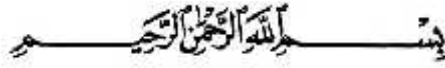
Informatics Engineering Departement  
Faculty of Science and Technology  
State Islamic University of Sultan SyarifKasim Riau  
Soebrantas Street No. 155 Pekanbaru

## **ABSTRACT**

The mining text is called the data mining text (TDM) or Knowledge Discovery in the Text is a new finding of information and it is known before by computer, with automatically to extracts information from the text resources, the text is not structure that different. The mining text system has known a method in a group of the text, one of them is clustering K-means. K-means is one of clustering method with partitioning approach to partition the existing data in the form of one or more group. The assessment clustering method used for winnowing algorithm, the function of algorithm to use to form document text can be a group of Fingerprint values. The K-means method use to Jaccard Coefficient and Euclidian Distance to measure the level of Similarity between document from result text to relevance jaccard coefficient get for 4 groups score relevance 73,5 %, from 7 groups 79 %, and for 5 groups 96,66 %, while relevance Euclidian distance get for 4 groups 49,5 %, 7 groups score relevance has failed, and 5 groups 58,33 %. The final result of this test get from Similarity method it is better from using Jaccard Coefficient.

**Key words:** :Clustering, Euclidian Distance, Jaccard Coefficient, K-means, Relevance, Similarity, Text Mining.

## KATA PENGANTAR



*Alhamdulillah Robbil' alamin*, penulis bersyukur ke-hadirat Allah SWT, karena atas segala limpahan rahmat dan karunia-Nya yang diberikan sehingga penulis dapat menyelesaikan penelitian dan penulisan laporan tugas akhir ini. *Allahumma sholli'ala Muhammad wa'ala ali sayyidina Muhammad*, yang tidak lupa penulis haturkan juga untuk Rosul Allah, Muhammad SAW.

Laporan tugas akhir ini merupakan salah satu prasyarat untuk memenuhi persyaratan akademis dalam rangka meraih gelar kesarjanaan di Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau (UIN SUSKA Riau). Selama menyelesaikan tugas akhir ini, penulis telah banyak mendapatkan bantuan, bimbingan, dan petunjuk dari banyak pihak baik secara langsung maupun tidak langsung. Untuk itu dalam kesempatan ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. H. M. Munzir Hitami, MA, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Ibu Dra. Hj. Yenita Morena, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Ibu Elin Haerani, ST, M.Kom, selaku Ketua Jurusan Teknik Informatika, Fakultas Sains dan Teknologi.
4. Bapak Drs. Martius, M.Hum, selaku Ketua sidang tugas akhir dan Wakil Dekan II Fakultas Sains dan Teknologi. Terimakasih pak.
5. Bapak Surya Agustian, ST, M.Kom selaku dosen pembimbing tugas akhir. Terimakasih pak untuk waktu yang selalu bapak luangkan untuk penulis, ilmu, semangat, dan motivasinya yang luar biasa. Terimakasih pak.
6. Bapak Jasril, S.Si, M.Sc, selaku dosen penguji 1 yang banyak membantu dan memberi masukan penulis dalam penyempurnaan Laporan Tugas Akhir ini, untuk ilmu-ilmunya.

7. Ibu Elin Haerani, ST, M.Kom, selaku dosen penguji 2, terimakasih ibu untuk ilmu-ilmunya, saran-sarannya, perbaikan-perbaikannya, dan masukannya.
  8. Bapak Muhammad Affandes, ST, MT, sebagai koordinator tugas akhir yang telah memberi masukan-masukan untuk penyelesaian tugas akhir ini, dan sangat sabar membantu penulis dalam mempersiapkan semua kebutuhan penulis dalam penyelesaian Tugas Akhir ini.
  9. Ayahanda tercinta Wisnu Rukminto, SE, yang menjadi sumber semangat penulis dan sudah mendidik penulis sehingga bisa menimba ilmu di jenjang perguruan tinggi seperti saat sekarang. Ibuanda tercinta Jumi Eliya, yang tiada hentinya memanjatkan doa, memberikan semangat untuk kesuksesan penulis.
  10. Indah Putri Ramadhan, seseorang yang istimewa yang dengan ikhlas dan tulus selalu ada memberikan semangat kepada penulis, dan adik penulis Mona Wiliyastari yang selalu memberikan bantuan dan masukan demi kesuksesan penulis.
  11. Teman-teman karib Undeka yang telah tulus membantu penulis selama ini, Sucipto, Arianda, Mbak Vera, Zulfan Effendi, Yuliska, Ersad, Ramon Muhandaz atas saran, bantuan dan do'a serta motivasi, terimakasih banyak.
  12. Teman-teman karib satu angkatan 2007 Teknik Informatika, Ihsan, Fairuz, Heri, Hallend, Hendra, dan terutamatemankelas B.
  13. Semua Pihak yang terlibat baik langsung maupun tidak langsung yang terlibat dalam pelaksanaan Tugas Akhir ini yang tidak dapat penulis sebutkan satu persatu
- Akhirnya, penulis menyadari dalam penulisan laporan ini masih terdapat kekurangan. Oleh karena itu, saran dan kritik sangat penulis harapkan untuk kemajuan penulis secara pribadi. Terimakasih.

Pekanbaru, 24 Juni 2014

Penulis

# DAFTAR ISI

	Halaman
HALAMAN JUDUL LAPORAN .....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN .....	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iv
LEMBAR PERNYATAAN.....	v
LEMBAR PERSEMBAHAN .....	vi
ABSTRAK.....	vii
<i>ABSTRACT</i> .....	viii
KATA PENGANTAR .....	ix
DAFTAR ISI .....	xi
DAFTAR GAMBAR .....	xiv
DAFTAR TABEL.....	xvi
DAFTAR LAMPIRAN.....	xviii
DAFTAR RUMUS .....	xix
DAFTAR SIMBOL.....	xx
BAB I PENDAHULUAN .....	I-1
1.1. Latar Belakang .....	I-1
1.2. Rumusan Masalah .....	I-2
1.3. Batasan Masalah.....	I-2
1.4. Tujuan Penelitian .....	I-2
1.5. Sistematika Penulisan .....	I-3
BAB II LANDASAN TEORI .....	II-1
2.1. Text Mining.....	II-1
2.2. Ruang Lingkup Text Mining.....	II-1
2.2.1. <i>Text Preprocessing</i> .....	II-2
2.2.2. <i>Text Transformation</i> .....	II-2
2.2.2.1. Algoritma Winnowing.....	II-3

2.2.3. <i>Feature Selection</i> .....	II-7
2.2.4. <i>Pattern Discovery</i> .....	II-7
2.2.4.1. Algoritma K-Means .....	II-8
2.2.5. <i>Clustering</i> .....	II-13
2.2.5.1. Metode <i>Clustering</i> .....	II-14
2.2.5.2. Klasifikasi <i>Clustering</i> .....	II-14
2.3. Metode Pengukuran Performansi .....	II-16
BAB III METODOLOGI PENELITIAN.....	III-1
3.1. Identifikasi Masalah .....	III-2
3.2. Perumusan Masalah.....	III-2
3.3. Studi Literatur .....	III-2
3.4. Analisa Sistem .....	III-2
3.5. Perancangan Aplikasi .....	III-4
3.6. Implementasi .....	III-4
3.7. Pengujian Aplikasi .....	III-5
3.8. Kesimpulan dan Saran.....	III-5
BAB IV ANALISIS DAN PERANCANGAN .....	IV-1
4.1. Analisa Dokumen Clustering .....	IV-1
4.2. Analisa <i>Teks Mining</i> .....	IV-2
4.2.1. <i>Text Preprocessing</i> .....	IV-4
4.2.2. <i>Text Transformation</i> .....	IV-6
4.2.3. <i>Feature Selection</i> .....	IV-14
4.2.4. <i>Pattern Discovery</i> .....	IV-17
4.3. Perancangan Aplikasi .....	IV-26
4.3.1. Perancangan Tampilan Menu.....	IV-26
4.3.2. Rancangan <i>Form Interface (Home)</i> .....	IV-27
4.3.3. Rancangan <i>Form Corpus</i> .....	IV-27
4.3.4. Rancangan <i>Form Setting</i> .....	IV-28
4.3.5. Rancangan <i>Form Clustering</i> .....	IV-29
4.3.6. Rancangan <i>Form Cek Gram</i> .....	IV-30
4.3.7. Rancangan <i>Form List Uji</i> .....	IV-31
4.3.8. Rancangan <i>Form About</i> .....	IV-31



BAB V IMPLEMENTASI DAN PENGUJIAN .....	V-1
5.1. Tahapan Implementasi .....	V-1
5.1.1. Batasan Implementasi .....	V-1
5.1.2. Lingkungan Implementasi .....	V-1
5.1.3. Implementasi Antarmuka Aplikasi .....	V-2
5.2. Pengujian Aplikasi .....	V-9
5.2.1. Rencana Pengujian .....	V-9
5.2.1.1 Pengujian Nilai Inputan Program .....	V-12
5.2.1.2 Pengujian Hasil Running Program .....	V-14
5.2.2. Hasil Pengujian .....	V-18
5.2.3. Pengujian <i>Blackbox</i> .....	V-19
5.2.4. Kesimpulan Pengujian .....	V-21
BAB VI PENUTUP .....	VI-1
6.1. Kesimpulan .....	VI-1
6.2. Saran .....	VI-2
DAFTAR PUSTAKA	
LAMPIRAN	
DAFTAR RIWAYAT HIDUP	