



**PENCARIAN KATA PADA DOKUMEN DENGAN CLUSTERING
MENGUNAKAN POLA VARIAN HILL CLIMBING**

SKRIPSI

Oleh

**TIARA KURNIA MAHISA
NIM. 091810101038**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2013**



**PENCARIAN KATA PADA DOKUMEN DENGAN CLUSTERING
MENGUNAKAN POLA VARIAN HILL CLIMBING**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan Program Studi Matematika (S1)
dan mencapai gelar Sarjana Sains

Oleh

**TIARA KURNIA MAHISA
NIM. 091810101038**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2013**

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Ibunda Ir. Tatik Amalijah dan Ayahanda Ir. Cucuk Haryandoko yang tercinta;
2. Adikku Nazwa Putri Nabila;
3. guru-guru sejak taman kanak-kanak sampai dengan perguruan tinggi;
4. Almamater Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

MOTO

Pendidikan merupakan perlengkapan paling baik untuk hari tua. *)

Genius without education is like silver in the mine. **)

*) Aristoteles. Kumpulan Motto Kehidupan [on line].
<http://pristality.wordpress.com/2011/02/23/kumpulan-motto-kehidupan/> [13 Mei 2013]

**) Benjamin Franklin .2010. *Temukan Penyebabnya Temukan Jawabannya*. Jakarta: PT Bestari Buana Murni

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

nama : Tiara Kurnia Mahisa

NIM : 091810101038

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Pencarian Kata pada Dokumen dengan Clustering menggunakan Pola Varian Hill Climbing” adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi mana pun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak mana pun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, September 2013

Yang menyatakan,

Tiara Kurnia Mahisa
NIM 091810101038

SKRIPSI

PENCARIAN KATA PADA DOKUMEN DENGAN CLUSTERING MENGUNAKAN POLA VARIAN HILL CLIMBING

Oleh

Tiara Kurnia Mahisa
NIM 091810101038

Pembimbing

Dosen Pembimbing Utama : Ahmad Kamsyakawuni, S.Si, M.Kom.

Dosen Pembimbing Anggota : Dr. Alfian Futuhul Hadi, S.Si, M.Si.

PENGESAHAN

Skripsi yang berjudul “Pencarian Kata pada Dokumen dengan Clustering menggunakan Pola Varian Hill Climbing” telah diuji dan disahkan pada:

hari, tanggal :

tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember

Tim Penguji:

Dosen Pembimbing Utama,

Dosen Pembimbing Anggota,

Ahmad Kamsyakawuni, S.Si, M.Kom.
NIP.197211291998021001

Dr. Alfian Futuhul Hadi, S.Si, M.Si.
NIP.197407192000121001

Penguji I,

Penguji II,

Yuliani Setia Dewi, S.Si, M.Si.
NIP.197407162000032001

Agustina Pradjaningsih, S.Si, M.Si.
NIP.197108022000032009

Mengesahkan

Dekan,

Prof. Drs. Kusno, DEA., Ph.D.
NIP 196101081986021001

RINGKASAN

Pencarian Kata pada Dokumen dengan Clustering menggunakan Pola Varian Hill Climbing. Tiara Kurnia Mahisa, 091810101038; 2013: 38 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Information Retrieval merupakan cabang ilmu yang membahas tentang pencarian informasi, baik berupa dokumen, referensi ilmiah, ataupun pengetahuan lainnya. Banyaknya informasi maupun referensi yang tidak disajikan secara digital menyulitkan pembaca untuk mencari secara manual. Selain itu, pembaca juga sulit untuk menemukan referensi yang memiliki pembahasan yang relevan dengan topik bacaan yang akan dicari. Tugas akhir ini mengaplikasikan algoritma *clustering* pada pencarian informasi menggunakan uji pola varian Hill Climbing, dan membuat program algoritma *clustering* dengan PHP dan MySQL.

Penelitian dilakukan dalam beberapa langkah, yaitu mengolah data berupa dokumen artikel ilmiah yang telah diperoleh, menjadi data penyimpanan dalam *database*. Langkah selanjutnya yaitu membuat algoritma pemrograman menggunakan algoritma *clustering* dan pengujian Hill Climbing, membuat program berdasarkan algoritma menggunakan bahasa pemrograman PHP dan MySQL, dan langkah terakhir yaitu menampilkan hasil dokumen yang dicari menggunakan program yang sudah dibuat.

Program aplikasi pencari kata yang telah dibuat mampu menampilkan dokumen yang dicari dan dokumen-dokumen yang mempunyai pembahasan yang terkait satu dengan yang lain. Dokumen yang dimunculkan pada *output* dengan mengaplikasikan pencarian kata menggunakan *clustering* tidak tergantung pada dokumen yang mengandung jumlah kata kunci paling banyak. Pemilihan *clustering*

yang terbaik dapat dicari dengan pengujian pola varian Hill Climbing, yaitu dengan memilih *clustering* yang mempunyai nilai varian terbesar. Salah satu hasil pencarian program tersebut adalah pencarian pada kata "penyebaran penyakit". Setelah dilakukan *clustering* pada kata tersebut, pemilihan *clustering* yang tepat setelah diuji dengan Hill Climbing adalah pada saat dibentuk sebanyak 3 *cluster*, karena pada pembentukan *cluster* tersebut menghasilkan nilai varian tertinggi yang berarti kedekatan dokumen dalam *cluster* tersebut sangat besar. Begitu juga dengan pencarian kata "pencarian jarak minimal" mempunyai nilai varian tertinggi saat dibentuk sebanyak 6 *cluster* setelah diuji dengan Hill Climbing.

PRAKATA

Puji syukur ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pencarian Kata pada Dokumen dengan Clustering menggunakan Pola Varian Hill Climbing”. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan strata satu (S1) pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penulisan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Allah SWT yang telah memberi petunjuk dan memberikan kekuatan yang luar biasa untuk menjalani hidup ini dengan penuh kesabaran dan keikhlasan;
2. Bapak Ahmad Kamsyakawuni, S.Si, M.Kom selaku Dosen Pembimbing Utama, dan Bapak Dr. Alfian Futuhul Hadi, S.Si, M.Si. selaku Dosen Pembimbing Anggota yang telah meluangkan waktu, pikiran, dan perhatian dalam penulisan skripsi ini;
3. Ibu Yuliani Setia Dewi, S.Si, M.Si. dan Ibu Agustina Pradjaningsih, S.Si, M.Si selaku dosen penguji yang telah memberi masukan dalam skripsi ini;
4. Bapak Ali Ridho Barakbah yang telah bersedia memberikan penjelasan materi tentang tugas akhir ini;
5. Ibu Dian Anggraini, S.Si dan Bapak Kusbudiono, S.Si, M.Si sebagai pembimbing akademik selama penulis menjadi mahasiswa Matematika MIPA;
6. Bapak Ziaul Arif, S.Si yang telah membantu penulis untuk mencari jurnal-jurnal yang berkaitan dengan tugas akhir ini;
7. Ibunda Ir. Tatik Amalijah, Ayahanda Ir. Cucuk Haryandoko, dan adik saya tersayang yang telah memberikan doa dan dorongannya demi terselesaikannya skripsi ini;

8. Raden Herda Widyatara yang sabar dan penuh pengertian dalam menemani serta mendukung segala usaha dalam penyelesaian tugas akhir ini ;
9. Dewi, Rini, Nila, Ranggi, Avan dan teman-teman MALINC, kakak serta adik angkatan Jurusan Matematika MIPA, dan semua pihak yang tidak dapat disebutkan satu per satu.

Penulis juga menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat.

Jember, September 2013

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	ii
HALAMAN PERSEMBAHAN	iii
HALAMAN MOTO	iv
HALAMAN PERNYATAAN	v
HALAMAN PEMBIMBINGAN	vi
HALAMAN PENGESAHAN	vii
RINGKASAN	viii
PRAKATA	x
DAFTAR ISI	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
BAB 2. TINJAUAN PUSTAKA	4
2.1 <i>Text Mining</i>	4
2.2 <i>Clustering</i>	6
2.2.1 <i>Hierarchical Clustering</i>	7
2.2.2 <i>Single Linkage Hierarchical Method (SLHM)</i>	8
2.2.3 <i>Analisis Cluster</i>	9
2.3 Hill Climbing	11

2.4 PHP dan MySQL	13
BAB 3. METODE PENELITIAN	15
BAB 4. HASIL DAN PEMBAHASAN	18
4.1 Hasil	18
4.1.1 Data	18
4.1.2 Penyelesaian <i>Text Mining</i> menggunakan Program	19
4.1.3 Langkah-langkah Menjalankan Program	19
4.2 Pembahasan.....	34
BAB 5. PENUTUP	36
5.1 Kesimpulan.....	36
5.2 Saran	36
DAFTAR PUSTAKA	37
LAMPIRAN	39

DAFTAR TABEL

	Halaman
Tabel 4.1	Jumlah Kata Kunci yang Ditemukan pada Tiap Dokumen dari kata ‘penyebaran penyakit’ 24
Tabel 4.2	Matriks Jarak antar Dokumen 25
Tabel 4.3	Tabel hasil perhitungan V_c^2 , V_w , V_b , dan V saat 4 <i>cluster</i> 26
Tabel 4.4	Tabel hasil perhitungan V_c^2 , V_w , V_b , dan V saat 3 <i>cluster</i> 26
Tabel 4.5	Matriks jarak antar dokumen setelah dibentuk 3 <i>cluster</i> 27
Tabel 4.6	Tabel hasil perhitungan V_c^2 , V_w , V_b , dan V saat 2 <i>cluster</i> 27
Tabel 4.7	Matriks jarak antar dokumen setelah dibentuk 2 <i>cluster</i> 27
Tabel 4.8	Tabel hasil perhitungan V_c^2 , V_w , V_b , dan V saat 1 <i>cluster</i> 28
Tabel 4.9	Tabel hasil perhitungan nilai beda tinggi pada kata “penyebaran penyakit” 28
Tabel 4.10	Jumlah Kata Kunci yang Ditemukan pada Tiap Dokumen dari kata “pencarian jarak minimal” 30
Tabel 4.11	Tabel hasil perhitungan nilai beda tinggi pada kata “pencarian jarak minimal” 32

DAFTAR GAMBAR

	Halaman
Gambar 2.1	Ilustrasi <i>Single Linkage Hierarchical Method</i> 9
Gambar 2.2	Pola Nilai Beda Hill Climbing 12
Gambar 3.1	Bagan Tahap Metode Penelitian 15
Gambar 4.1	Tampilan Awal Program 20
Gambar 4.2	Tampilan Menu Utama untuk Admin 21
Gambar 4.3	Tampilan Program Saat <i>User</i> Memasukkan Kata Kunci 22
Gambar 4.4	<i>Output</i> dari Kata Kunci “penyebaran penyakit” Saat Nilai $\alpha=2$ 23
Gambar 4.5	Ilustrasi letak dokumen dalam diagram <i>cartesius</i> 25
Gambar 4.6	Grafik Perhitungan Uji Beda Tinggi pada Tiap <i>Cluster</i> Saat $\alpha=2$ pada Kata “penyebaran penyakit” 29
Gambar 4.7	Grafik Perhitungan Uji Beda Tinggi pada Kata “pencarian jarak minimal” 33
Gambar 4.8	Hasil <i>Output</i> dari Kata Kunci “pencarian jarak minimal” 34

DAFTAR LAMPIRAN

	Halaman
A. Hasil Perhitungan jarak antar dokumen pada kata “pencarian jarak minimal”	39
B. Hasil Perhitungan V_c^2 , V_w , V_b , dan V tiap Pembentukan <i>Cluster</i> pada Kata “pencarian jarak minimal”	40
C. <i>Flowchart</i> Algoritma <i>Single Linkage Hierarchical Method</i> dan Hill Climbing	44
D. Skrip Program Algoritma <i>Single Linkage Hierarchical Method</i> dan Hill Climbing	49