

IMPLEMENTASI ALGORITMA SINGLE LINKAGE UNTUK CLUSTERING DOKUMEN TEKS

Tugas Akhir



oleh

Vika Serlita
22064075

Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Tahun 2012

IMPLEMENTASI ALGORITMA SINGLE LINKAGE UNTUK CLUSTERING DOKUMEN TEKS

Tugas Akhir



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh:

Vika Serlita
22064075

Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Tahun 2012

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul :

IMPLEMENTASI ALGORITMA SINGLE LINKAGE UNTUK CLUSTERING DOKUMEN TEKS

Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan sarjana Program Studi Teknik Informatika, Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian sumber informasinya dicantumkan sebagaimana mestinya.

Jika kemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 4 Juni 2012



(Vika Serlita)

22064075



HALAMAN PERSETUJUAN

Judul : Implementasi Algoritma Single Linkage Untuk
Clustering Dokumen Teks
Nama : Vika Serlita
NIM : 22064075
Mata Kuliah : Tugas Akhir
Kode : TIW276
Semester : Genap
Tahun Akademik : 2011/2012

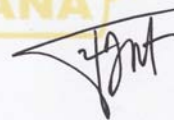
Telah diperiksa dan disetujui
Di Yogyakarta,
Pada tanggal 4 Juni 2012

Dosen Pembimbing I



Lucia Dwi Krisnawati, M.A.

Dosen Pembimbing II



Antonius Rachmat C, S.Kom., M.Cs.

HALAMAN PENGESAHAN

SKRIPSI
IMPLEMENTASI ALGORITMA SINGLE LINKAGE UNTUK
CLUSTERING DOKUMEN TEKS

Oleh : Vika Serlita / 22064075


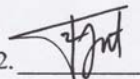
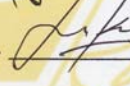
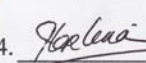
Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknik Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal
22 Juni 2012

Yogyakarta, 3 Juli 2012

Mengesahkan,

Dewan Penguji :


1. Lucia Dwi Krisnawati, M.A.
2. Antonius Rachmat C., S.Kom., M.Cs.
3. Lukas Chrisantyo, M.Eng.
4. Theresia Herlina R., S.Kom., M.T.

1. 
2. 
3. 
4. 

Dekan


(Drs. Wimmie Handiwidjojo, M.IT.)

Ketua Program Studi


(Nugroho Agus. H, S.Si, M.Si)

UCAPAN TERIMA KASIH

Terima kasih kepada Tuhan Yang Maha Esa karena telah mengizinkan dan membimbing penulis dalam menyusun dan menyelesaikan Tugas akhir dengan judul Implementasi Algoritma Single Linkage Untuk Clustering Dokumen Teks.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunanya.

Selama pembuatan Tugas Akhir ini, mulai dari perencanaan awal, penyusunan hingga ke penyelesaian baik program maupun laporan, telah memberi banyak sekali pengalaman dan pelajaran yang sangat berharga bagi penulis. Tentunya itu semua tidak lepas dari bimbingan dan dukungan dari banyak pihak. Dan pada kesempatan ini, penulis ingin sekali mengucapkan terima kasihnya kepada

1. Kedua Orang Tua penulis yang selalu memberikan dukungan dan doa-doa terbaik mereka kepada penulis dalam menyelesaikan pembuatan Tugas Akhir ini.
2. Ibu Lucia Dwi K, S.S, M.A, selaku dosen pembimbing I yang telah membimbing dengan sabar dari awal sampai akhir pembuatan Tugas Akhir ini.
3. Bapak Antonius Rachmat, S.Kom, M.Cs, selaku dosen pembimbing II yang telah membimbing dengan sabar dari awal sampai akhir pembuatan Tugas Akhir ini.
4. Saudara-saudara penulis, yang selalu memberikan dukungan moral dan “materi” ketika penulis mengalami masa-masa sulit dalam pengerjaan Tugas Akhir ini.

5. Teman baik yang selalu ada dan mendengarkan, Nana & Ci Vely yang telah tulus membantu dan selalu menemani penulis dalam pengerjaan Tugas Akhir ini.
6. Dan pihak-pihak lainnya yang tidak dapat penulis sebutkan satu per satu dalam pengerjaan Tugas Akhir ini.

Penulis sangat menyadari masih banyaknya kekurangan baik dalam program dan laporan Tugas Akhir ini. Sehingga, kritik dan saran sangat diharapkan oleh penulis dari para pembaca. Akhir kata, penulis berharap agar program dan laporan Tugas Akhir ini dapat bermanfaat bagi semua pihak yang membacanya.

Yogyakarta, Juni 2012

Vika Serlita



INTISARI

Kemudahan dalam mendapatkan informasi di jaman sekarang membuat sering sekali terjadi penimbunan di tempat penyimpanan. Informasi yang penting maupun tidak penting bercampur menjadi satu sehingga sulit untuk membedakan mana informasi yang benar-benar dibutuhkan dan tidak. Sangat penting adanya proses pengelompokan terhadap informasi-informasi tersebut. Pengelompokan ini dapat dilakukan secara digital(otomatis) maupun manual. Secara manual, tidak dapat berlangsung optimal, disebabkan kesulitan dalam mengelompokkan ataupun penentuan topik untuk tiap dokumen yang bisa sampai ribuan jumlahnya. Yang dapat memakan banyak waktu, tenaga, biaya dan tidak menutup kemungkinan terjadinya kesalahan dalam proses tersebut. Sehingga, diperlukan otomatisasi untuk mengelompokkan dokumen teks yang mempermudah user dalam mencari dan menganalisa suatu informasi.

Clustering Hierarki menggunakan algoritma Single Linkage merupakan solusi untuk membantu proses pengelompokan. Proses pengelompokan dimulai dari pembersihan isi dokumen. Dilanjutkan dengan mencari nilai kemiripan antar tiap dua dokumen. Dua dokumen dengan nilai kemiripan yang paling tinggi akan bergabung membentuk satu cluster baru. Proses penggabungan ini berlangsung sampai terbentuk satu cluster tunggal yang berisi semua dokumen. Proses penggabungan yang terjadi di tiap level menghasilkan bentuk hierarkis dari cluster.

Untuk mengukur nilai penyebaran data-data dari hasil clustering, dihitung nilai *variance within cluster*(V_w), nilai *variance between cluster*(V_b), dan nilai *variance*(V). Sistem menghasilkan nilai V_w sebanyak 57.14% lebih kecil dari proses manual, nilai V_b sebanyak 85.71% lebih besar dari proses manual, dan nilai V sebanyak 57.14% lebih kecil dari proses manual. Proses manual menghasilkan nilai

V_w sebanyak 28.57% lebih kecil dari sistem, nilai V_b sebanyak 14.29% lebih besar dari sistem, dan nilai V sebanyak 14.29% lebih kecil dari sistem. Antara sistem dan proses manual juga menghasilkan nilai yang sama untuk V_w sebanyak 14.29% , nilai V sebanyak 28.57%, dan tidak pernah bernilai sama untuk V_b . Dari hasil perhitungan nilai varian terlihat sistem menghasilkan nilai *variance* yang lebih kecil dari proses manual, yang berarti hasil clustering sistem lebih ideal/berkualitas lebih baik dari hasil clustering proses manual.

© UKDW

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN SKRIPSI.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
UCAPAN TERIMA KASIH.....	iv
INTISARI.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB 1 PENDAHULUAN	
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Hipotesis.....	3
1.5 Tujuan Penulisan.....	3
1.6 Metode Penelitian.....	3
1.7 Sistematika Penulisan.....	4
BAB 2 LANDASAN TEORI	
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	7
1. Hierarchical Clustering.....	7

2. Single Linkage.....	9	
3. Cosine Similarity Measure.....	10	
4. Variance (Analisis Cluster).....	14	10

BAB 3 PERANCANGAN SISTEM

3.1 Alat dan Bahan Penelitian.....	16	
3.1.1 Alat Penelitian.....	16	
3.1.2 Bahan Penelitian.....	16	
3.2 Use Case Diagram.....	17	
3.3 Activity Diagram.....	19	
3.4 Perancangan Masukan.....	20	
3.4.1 Masukan Dokumen.....	20	
3.4.2 Masukan Stopword.....	20	
3.5 Perancangan Luaran.....	20	
3.6 Rancangan Proses Sistem Clustering.....	21	
3.7 Rancangan Proses Input dan Data Cleaning.....	22	
3.8 Rancangan Proses Pembobotan Dokumen.....	23	
3.9 Rancangan Proses Clustering.....	23	
3.10 Contoh Perhitungan Manual Sistem.....	24	
3.11 Rancangan Database.....	32	
3.12 Rancangan Antarmuka.....	33	
3.12.1 Rancangan Form Input Dokumen.....	33	
3.12.2 Rancangan Form Tampil Hasil Clustering.....	34	
3.12.3 Rancangan Form Tampil Data Bersih.....	35	
3.12.4 Rancangan Form Tampil Bobot Token.....	36	

3.12.5 Rancangan Form Tampil Nilai Similarity.....	36
3.12.6 Rancangan Form Stopword.....	37
3.13 Rancangan Pengujian.....	37
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM	
4.1 Implementasi Sistem.....	38
4.1.1 Implementasi <i>User Interface</i>	38
4.1.1.1 Implementasi Tampilan Input.....	38
4.1.1.1.1 Implementasi Form Utama.....	38
4.1.1.1.2 Implementasi Form Stopword.....	39
4.1.1.2 Implementasi Tampilan Output.....	40
4.1.1.2.1 Implementasi Tampilan Data Bersih.....	40
4.1.1.2.2 Implementasi Tampilan Bobot Token.....	41
4.1.1.2.3 Implementasi Tampilan Nilai Similarity.....	42
4.1.1.2.4 Implementasi Tampilan Hasil Clustering.....	42
4.1.1.3 Implementasi Tampilan Tambahan.....	43
4.1.1.3.1 Implementasi Tampilan Form Bantuan.....	44
4.1.1.3.2 Implementasi Tampilan Form Tentang Sistem.....	44
4.1.2 Implementasi Proses.....	45
4.1.2.1 Implementasi Proses Input Dokumen.....	45
4.1.2.2 Implementasi Proses Pembobotan.....	47
4.1.2.3 Implementasi Proses Menghitung Nilai Similarity.....	49
4.1.2.4 Implementasi Proses Clustering.....	50
4.2 Analisis Sistem.....	51
4.2.1 Analisis Nilai Precision dan Purity.....	51

4.2.2 Analisis Tingkatan Hirarki Clustering.....	63
BAB 5 KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	81
5.2 Saran.....	81
DAFTAR PUSTAKA.....	82
LAMPIRAN A (<i>source code</i>)	A-1

DAFTAR TABEL

Tabel 3.1 Tabel Df.....	26
Tabel 3.2 Tabel TF.....	27
Tabel 3.3 Tabel IDF.....	27
Tabel 3.4 Tabel $w_{(i,j)}$	28
Tabel 4.1 Tabel Hasil Clustering Kasus 1.....	52
Tabel 4.2 Tabel Hasil Clustering Kasus 2.....	53
Tabel 4.3 Tabel Hasil Clustering Kasus 3.....	54
Tabel 4.4 Tabel Hasil Clustering Kasus 4.....	56
Tabel 4.5 Tabel Hasil Clustering Kasus 5.....	57
Tabel 4.6 Tabel Hasil Clustering Kasus 6.....	59
Tabel 4.7 Tabel Hasil Clustering Kasus 7.....	61

DAFTAR GAMBAR

Gambar 2.1 Proses Pembentukan Cluster	8
Gambar 3.1 Use Case Diagram Sistem Clustering.....	17
Gambar 3.2 Activity Diagram Input Dokumen.....	19
Gambar 3.3 Activity Diagram Clustering.....	19
Gambar 3.4 Flowchart Sistem Clustering Secara Keseluruhan.....	21
Gambar 3.5 Flowchart Proses Input dan Data Cleaning.....	22
Gambar 3.6 Flowchart Rancangan Proses Pembobotan.....	23
Gambar 3.7 Flowchart Rancangan Proses Clustering.....	24
Gambar 3.8 Matriks Similarity Antar Dokumen.....	30
Gambar 3.9 Matriks Nilai Similarity D1 dan D3.....	30
Gambar 3.10 Matriks Nilai Similarity D13 dan D2.....	31
Gambar 3.11 Matriks Nilai Similarity D132 dan D4.....	31
Gambar 3.12 Rancangan Database Sistem Clustering.....	33
Gambar 3.13 Rancangan Form Input Dokumen.....	34
Gambar 3.14 Rancangan Form Tampil Hasil Clustering.....	35
Gambar 3.15 Rancangan Form Tampil Data Bersih.....	35
Gambar 3.16 Rancangan Form Tampil Bobot Token.....	36
Gambar 3.17 Rancangan Form Tampil Nilai Similarity.....	36
Gambar 3.18 Rancangan Form Stopword.....	37
Gambar 4.1 Form Utama Sistem Clustering.....	39
Gambar 4.2 Form Stopword.....	40
Gambar 4.3 Form Tampil Data Bersih.....	41

Gambar 4.4 Form Tampil Bobot Token.....	41
Gambar 4.5 Form Tampil Nilai Similairity.....	42
Gambar 4.6 Form Tampil Hasil Clustering.....	43
Gambar 4.7 Form Bantuan.....	44
Gambar 4.8 Form Tentang Sistem.....	44
Gambar 4.9 Pseudoce Input Dokumen.....	46
Gambar 4.10 Pseudoce Data Cleaning().....	46
Gambar 4.11 Pseudoce Tokenisasi().....	46
Gambar 4.12 Pseudoce Stopword().....	47
Gambar 4.13 Pseudoce Lexicon().....	48
Gambar 4.14 Pseudoce TF-Idf.....	49
Gambar 4.15 Pseudoce Hitung Nilai Similarity.....	50
Gambar 4.16 Pseudoce Proses Clustering.....	51



BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi komputer khususnya di jaringan Internet telah berkembang dengan sangat cepat. Semua orang dapat saling bertukar dan mendapatkan informasi melalui jaringan Internet. Informasi-informasi berupa dokumen teks, gambar, maupun audio dapat dengan mudah diakses melalui jaringan Internet dan kemudian diunduh. Dengan segala kemudahan tersebut, semua informasi dapat diperoleh, diunduh, dan kemudian disimpan dalam media penyimpanan komputer. Terlebih semua orang dapat membuat, menulis, dan menyimpan sebuah informasi. Hal inilah yang secara tidak langsung menimbulkan penimbunan informasi. Informasi yang penting maupun tidak penting bercampur menjadi satu sehingga sulit untuk membedakan mana informasi yang benar-benar kita butuhkan dan tidak.

Sangat penting adanya proses seleksi terhadap informasi-informasi tersebut. Penyeleksian ini dapat dilakukan secara digital (otomatis) maupun manual. Secara manual, penyeleksian tidak dapat berlangsung optimal. Hal ini disebabkan kesulitan dalam melakukan pengelompokan informasi ataupun penentuan topik secara manual untuk tiap dokumen yang bisa saja sampai ribuan jumlahnya, yang mana akan memakan banyak waktu, tenaga, biaya dan bahkan tidak menutup kemungkinan terjadinya kesalahan dalam proses seleksi tersebut. Sehingga, diperlukan otomatisasi untuk pengelompokan informasi teks yang tidak sedikit jumlahnya, dan untuk mempermudah user dalam mencari dan menganalisa suatu informasi.

Otomatisasi yang diharapkan menggunakan teknik clustering dalam melakukan pengelompokan dokumen-dokumen teks yang telah ada. Clustering dokumen teks adalah proses menentukan cluster suatu dokumen teks tertentu.

Clustering dokumen teks ini menentukan kelompok suatu teks, termasuk cluster manakah suatu dokumen teks tertentu.

1.2 Perumusan Masalah

Dari latar belakang yang telah dikemukakan di atas, maka akan dirumuskan beberapa masalah yaitu :

1. Bagaimana penerapan metode *Single Linkage* untuk meng-*cluster* -kan sekumpulan dokumen teks yang sudah ada?
2. Bagaimana analisa hasil *clustering* dokumen dengan metode *Single Linkage*?

1.3 Batasan Masalah

Program clustering yang akan dikembangkan penulis memiliki beberapa batasan :

1. Program yang dibuat tidak dapat diimplementasikan ke dalam jaringan komputer.
2. Dokumen text yang dapat diproses adalah *file plain text* yang menggunakan bahasa Indonesia. (berektensi .txt)
3. Jumlah maksimal *file .txt* yang diteliti adalah 200 file.
4. Dokumen diambil dari <http://www.kompas.com/> sebanyak 200 buah, dengan 50 buah dokumen Bisnis, 50 buah dokumen Edukasi, 50 buah dokumen Olahraga, dan 50 buah dokumen Travel.
5. Tidak dilakukan proses *stemming* untuk setiap token.
6. Tidak menerima inputan *query*.

1.4 Hipotesis

Dari penelitian ini, penulis membuat dugaan bahwa program clustering yang akan dibuat dengan menggunakan metode Single Linkage mampu mengclusterkan dokumen-dokumen sesuai dengan tingkat kemiripannya. Sehingga dapat terlihat bentuk hierarki dari dokumen-dokumen yang telah mengalami proses clustering.

1.5 Tujuan Penelitian

Tujuan yang ingin dicapai dari usulan penulisan skripsi ini adalah :

1. Menghasilkan sebuah program bantu yang dapat digunakan untuk meng- *cluster* -kan dokumen teks.
2. Menguji dan mengimplementasikan secara langsung algoritma Single Linkage dan Tf-Idf yang akan dipakai, untuk melihat bentuk hierarki dari dokumen-dokumen yang telah mengalami proses pengclusteran.

1.6 Metode/Pendekatan

Berikut ini adalah metodologi - metodologi yang akan digunakan dalam melaksanakan penelitian ini :

- **Metode Pengumpulan Data**
 - **Studi Pustaka**

Penulis melakukan studi pustaka yang dilakukan dengan cara mempelajari teori-teori dan literatur-literatur yang mendukung penyelesaian penelitian ini, terutama yang berhubungan dengan algoritma *Single Linkage* dan perangkat lunak yang digunakan untuk membangun sistem.

- **Pembuatan Korpus**

Penulis mengumpulkan *file* .txt dari <http://www.kompas.com/> sebagai bahan untuk korpus data yang digunakan dalam proses penelitian.

- **Metode Pengembangan Sistem**

Metode yang dipakai dalam pengembangan sistem dalam penelitian ini adalah metode *Single Linkage* yang merupakan bagian dari *hierarchical clustering*.

- **Metode Evaluasi**

Metode evaluasi yang akan digunakan adalah melakukan perbandingan antara hasil *clustering* keluaran dari sistem dan hasil *clustering* secara manual oleh penulis dengan cara menganalisa tiap dokumen.

1.7 Sistematika Penulisan

Adapun sistematika laporan tugas akhir ini dibagi menjadi 5 bab besar, yaitu :

Bab 1 Pendahuluan berisi latar belakang masalah, perumusan masalah, batasan masalah, hipotesis, tujuan penelitian, metode penelitian dan sistematika penulisan laporan.

Bab 2 Tinjauan pustaka yang berisi dasar teori yang diperlukan dalam penulisan tugas akhir ini seperti teori tentang *hierarchical clustering*, *single linkage clustering* dan lain-lain.

Bab 3 Perancangan Sistem membahas tentang bagaimana sistem yang akan dibangun meliputi alur kerja sistem, antar muka sistem, dan perancangan basis data pada sistem.

Bab 4 Implementasi dan Analisis Sistem yang akan membahas tentang metode penelitian yang digunakan beserta langkah-langkah serta hasil dari penelitian. Dari bab ini, didapatkan kelebihan dan kekurangan sistem.

Bab 5 Kesimpulan dan Saran merupakan bagian terakhir dari karya tulis ini, dimana berisi beberapa intisari penting yang dapat diambil dari penelitian dan juga saran pengembangan sistem bagi penelitian berikutnya.

© UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, penulis mengambil kesimpulan bahwa :

1. Algoritma Single Linkage berhasil diterapkan pada sistem, dan mampu menghasilkan *cluster-cluster* dengan isi dokumen yang mirip.
2. Kualitas dari hasil *clustering* tidak tergantung oleh jumlah korpus dan jumlah kategori, karena yang menjadi acuan adalah nilai kemiripan dari isi dokumen.
3. Dari hasil perhitungan Varian disimpulkan bahwa, semakin besar nilai varian pada suatu level, maka semakin rendah kualitas dari *cluster* di level tersebut. Sedangkan, semakin kecil nilai varian pada suatu level maka semakin tinggi kualitas dari *cluster* di level tersebut.

5.2 Saran

Saran untuk pengembangan sistem lebih lanjut adalah :

1. Diutamakan penyimpanan utama pada proses *clustering* menggunakan *array*. Dengan pertimbangan dapat menghemat waktu komputasi yang lebih banyak.
2. Menambahkan perhitungan V_b , V_w , dan Varian semua *cluster* ke dalam sistem, agar pengguna dapat mengetahui secara langsung kualitas dari hasil *cluster* yang dihasilkan.

Daftar Pustaka

- Billsus, D., & Pazzani, M.J. (1998) Learning Collaborative Information Filters. *Proceedings of the Fifteenth International Conference on Machine Learning*, 46-54. Morgan Kaufmann Publisher Inc. San Francisco, CA, USA.
- El-Hamdouchi, A., & P. Willet. (1987). Comparison of Hierarchic Agglomerative Clustering Methods for Document Retrieval. *The Computer Journal*. Vol. 32 (3). Diakses pada tanggal 21 September 2010, dari <http://comjnl.oxfordjournals.org/content/32/3/220.full.pdf#page=1&view=FitH>
- Fowler, M. (2004). Uml Distilled 3th Ed., Panduan Singkat Bahasa Pemodelan Objek Standar. Yogyakarta: Penerbit Andi.
- Gregorius S.B., Rahardjo, A.I., & Taufik, H. (2008). Hierarchical Clustering Untuk Aplikasi Automated Text Integration. *Seminar Nasional Aplikasi Teknologi Informasi*. Yogyakarta : SNATI.
- Hamzah, A., Susanto, A., Soesianto, F., & Istiyanto, E. (2007) . Studi Komparasi Algoritma Hierarchical Dan Partitional Untuk Clustering Dokumen Teks Berbahasa Indonesia. *Jurnal Teknologi Academia Ista*. Vol. 12 (1).
- Hartini, E. (2004). *Metode Clustering Hirarki*. Diakses 18 September 2010, dari http://www.batan.go.id/ppin/lokakarya/LKSTN_15/Entin.pdf
- Intan, R., & Defeng, A., (2006). Hard: Subject-Based Search Engine Menggunakan TF-IDF Dan Jaccard's Coefficient. *Jurnal Teknik Industri*. Vol. 8 (1), 61-72.
- Primadhanty, R.A. (2007). *Studi dan Implementasi Active Fuzzy Constrained Clustering untuk Pengelompokan Dokumen*. Diakses 15 September 2010, dari <http://www.digilib.itb.ac.id>

Rosell, M. (2009). *Introduction to Information Retrieval and Text Clustering*. Diakses 18 September 2010, dari

<http://www.nada.kth.se/~rosell/undervisning/sprakt/irintro090824.pdf>

Steinbach, M., Karypis, G., & Kumar, V. (2008). *A Comparison of Document Clustering Techniques*. Diakses 21 September 2010, dari

<http://en.scientificcommons.org/42586635>

Qian, G., Sural, S., Gu, Y., & Pramanik, S. (2004). *Similarity between Euclidean and Cosine Angle Distance for Nearest Neighbor Queries*. Diakses 14 September

2011, dari <http://www.citeseerx.ist.psu.edu>

© UKDW