

**PENERAPAN K-MEDOIDS DALAM PENGELOMPOKAN  
SKRIPSI MAHASISWA**

**Studi Kasus Mahasiswa Fakultas Teknologi Sistem Informasi UKDW**

**Universitas Kristen Duta Wacana**

**Skripsi**



Disusun oleh:

Elvira Ratna Aghasty

23080436

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
TAHUN 2011/2012**

**PENERAPAN K-MEDOIDS DALAM PENGELOMPOKAN  
SKRIPSI MAHASISWA**

**Studi Kasus Mahasiswa Fakultas Teknologi Sistem Informasi UKDW  
Universitas Kristen Duta Wacana**

**Skripsi**



Diajukan kepada Fakultas Teknologi Informasi Sistem informasi  
Universitas Kristen Duta Wacana

Sebagai salah satu syarat dalam memperoleh gelar  
Sarjana Komputer

Disusun oleh:

Elvira Ratna Aghasty

23080436

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
TAHUN 2011/2012**

## PERNYATAAN KEASLIAN KARYA ILMIAH SKRIPSI

Dengan ini saya,

Nama : Elvira Ratna Aghasty

No. Induk Mahasiswa : 23080436

Adalah mahasiswa **Fakultas Teknologi Informasi Program Studi Sistem Informasi, Universitas Kristen Duta Wacana** yang melakukan penelitian dengan :

**Judul Penelitian : Penerapan K-Medoids dalam Pengelompokan Skripsi Mahasiswa Studi Kasus : Mahasiswa Fakultas Teknologi Informasi UKDW**

menyatakan dengan sungguh-sungguh dan benar bahwa dalam mengerjakan penelitian dengan judul di atas saya :

1. tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan
2. tidak melakukan plagiat terhadap naskah karya orang lain
3. tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa ijin pemilik karya
4. tidak melakukan manipulasi dan pemalsuan data
5. mengerjakan sendiri karya ini dan mampu bertanggung jawab atas karya ini

Jikalau di kemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggung-jawabkan, ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan di atas, maka saya siap untuk dikenai sanksi apapun termasuk pembatalan nilai tugas akhir saya oleh Universitas Kristen Duta Wacana. Demikian pernyataan ini saya buat dengan sesungguhnya.

Yogyakarta, 12 Desember 2011

Yang menyatakan,



(Elvira Ratna Aghasty)

## HALAMAN PERSETUJUAN

Judul : Penerapan K-Medoids dalam Pengelompokan Skripsi  
Mahasiswa Studi Kasus : Mahasiswa Fakultas Teknologi  
Informasi UKDW

Nama : Elvira Ratna Aghasty

NIM : 23080436

Mata Kuliah : Skripsi

Kode Mata Kuliah : SI4046

Semester : Ganjil

Tahun : 2011/2012

Fakultas : Teknologi Informasi

Program Studi : Sistem Informasi

---



Dosen Pembimbing I,

Umi Probayekti, S.Kom, MLIS.

Dosen Pembimbing II,

Drs. Jong Jek Siang, M.Sc.

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PENERAPAN K-MEDOIDS DALAM PENGELOMPOKAN**

**SKRIPSI MAHASISWA**

*Studi Kasus : Mahasiswa Fakultas Teknologi Informasi UKDW*

Oleh : Elvira Ratna Aghasty / 23080436

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Sistem Informasi Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
dan dinyatakan diterima untuk memenuhi salah satu  
syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal  
21 Desember 2011

Yogyakarta, 16 Januari 2012

Mengesahkan,

Dewan Penguji:

- 1) Umi Proboyekti, S.Kom, MLIS.
- 2) Drs. Jong Jek Siang, M.Sc.
- 3) Ir. Harianto Kristanto, M.T., M.M.
- 4) Lucia Dwi Krisnawati, S.S., M.A.

Dekan,

  
(Drs. Wimmie Handiwidjojo., MIT.)

Ketua Program Studi,

  
(Yetli Oslan, S.Kom, M.T.)

## UCAPAN TERIMA KASIH / KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus atas berkat-Nya sehingga penyusun dapat menyelesaikan Tugas Akhir dengan baik dan tepat waktu. Tugas Akhir berjudul Penerapan K-Medoids dalam Pengelompokan Skripsi Mahasiswa Studi Kasus : Mahasiswa Fakultas Teknologi Informasi UKDW dikerjakan sebagai salah satu syarat akademik guna memperoleh gelar Sarjana Strata Satu Universitas Kristen Duta Wacana Yogyakarta dan sebagai metode pembelajaran untuk menghasilkan sebuah karya yang dapat berguna.

Dalam kesempatan ini, penyusun ingin mengucapkan terima kasih kepada banyak pihak yang telah memberikan bimbingan, bantuan, doa, dan dukungan, sehingga Tugas Akhir ini dapat dikerjakan dan diselesaikan dengan baik dan tepat waktu, yaitu kepada :

- 1) Ibu Umi Proboyekti, S.Kom, MLIS., selaku Dosen Pembimbing I yang telah banyak memberikan perhatian, bimbingan, saran, dan bantuan dalam proses pengerjaan Tugas Akhir.
- 2) Bapak Drs. Jong Jek Siang, M.Sc., selaku Dosen Pembimbing II yang memberikan saran untuk Tugas Akhir dan yang selalu mengingatkan mengenai waktu pengerjaan Tugas Akhir.
- 3) Bapak Budi Susanto, S.Kom, M.T., selaku dosen yang mau memberikan kesempatan untuk belajar mengenai metode K-Medoids dan *Text Mining*, memberikan arahan dalam kendala yang ditemui pada proses pengerjaan Tugas Akhir, dan memberikan ijin untuk boleh menggunakan kamus dan daftar *stopword* yang mendukung pengerjaan Tugas Akhir.
- 4) Papa, mama, dan adik, yang selalu memberikan dukungan, semangat, doa, kasih sayang, dan perhatian sebagai keluarga.
- 5) Bernadetha Ayu yang bersedia membantu dalam memahami mengenai *web service* SInTA dan kendala yang ditemui dalam pengerjaan Tugas Akhir.

- 6) Simon Megadewandanu yang telah membantu dalam proses pembelajaran *framework* yang dipakai pada Tugas Akhir.
- 7) Teman-teman seperjuangan yang telah memberikan semangat, doa, kritik, dan saran, selama proses pengerjaan Tugas Akhir.
- 8) Semua pihak yang tidak dapat penyusun sebutkan satu-satu yang membantu penyusun baik secara langsung maupun tidak langsung.

Penyusun menyadari masih banyak kekurangan pada Tugas Akhir yang dikerjakan, oleh karena itu, kritik dan saran akan sangat bermanfaat bagi penyusun untuk dapat mengembangkan karya dari Tugas Akhir, sehingga menjadi lebih baik.

Akhir kata, penyusun berharap semoga karya yang dihasilkan dari Tugas Akhir ini dapat bermanfaat bagi pihak yang menggunakan.

Yogyakarta, 11 Januari 2012

Penyusun



**MOTTO**

*Lakukan yang terbaik yang kamu bisa, lakukan untuk TUHAN*

© UKDW



## ABSTRAK

SInTA (Sistem Informasi Tugas Akhir) UKDW merupakan sistem informasi yang menyimpan dan menampilkan kumpulan dokumen skripsi dari mahasiswa UKDW Yogyakarta. Dosen dan mahasiswa bahkan masyarakat diluar UKDW mendapatkan hak akses yang sama untuk melihat kumpulan skripsi yang tersedia. Masalah timbul ketika dosen ingin mengembangkan topik yang jarang atau paling sedikit dikerjakan untuk skripsi. Dosen memiliki kebutuhan untuk mengetahui kelompok topik skripsi yang paling banyak atau paling sedikit yang telah dibimbing, hanya SInTA tidak menyediakan fasilitas untuk menyajikan pengelompokan skripsi tersebut.

Dalam skripsi ini dibuat program untuk mengelompokkan skripsi bimbingan dosen dengan menerapkan metode K-Medoids.. Sistem melakukan pengelompokan dengan cara *clustering*. *Clustering* merupakan salah satu cara pengelompokan berdasarkan kesamaan karakteristik antar dokumen skripsi. Hasil pengelompokan didapatkan setelah pengguna memberi masukan berupa nama dosen, jumlah kelompok dan prosentase kemiripan dokumen pada sistem. Sistem akan menampilkan hasil berupa data skripsi yang telah dikelompokkan, nilai *purity* dari *clustering*, informasi dasar tentang jumlah dokumen, nilai *feature selection* dan nama dosen yang dipilih.

Hasil dari penelitian ini adalah program bantu pengelompokkan skripsi mahasiswa sesuai dengan kesamaan jenis dokumen berdasarkan abstrak skripsi. Metode K-Medoids mengelompokkan skripsi mahasiswa dengan menentukan medoid setiap *cluster* dan menghitung *Total Distance* paling kecil dari kombinasi medoid *cluster* yang dihasilkan. Pengelompokkan dokumen menghasilkan *purity* yang tinggi apabila jumlah kelompok yang ditentukan semakin besar dan keragaman jenis dokumen skripsi tergolong homogen. Nilai *purity* dari hasil *clustering* dipengaruhi oleh jumlah kelompok.

## DAFTAR ISI

HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
UCAPAN TERIMA KASIH / KATA PENGANTAR.....	v
MOTTO.....	vii
ABSTRAK.....	viii
DAFTAR ISI.....	ix
BAB 1.....	1
PENDAHULUAN.....	1
1.1    Latar Belakang Masalah.....	1
1.2    Perumusan Masalah.....	1
1.3    Batasan Masalah.....	2
1.4    Tujuan Penelitian.....	2
1.5    Metode / Pendekatan.....	2
1.6    Sistematika Penulisan.....	3
BAB 2.....	5
TINJAUAN PUSTAKA.....	5
2.1.    Tinjauan Pustaka.....	5
2.2.    Landasan Teori.....	6
2.2.1.    Text Mining.....	6
2.2.2.    Term Frequency-Invers Document Frequency (TF-IDF).....	8
2.2.3.    Algoritma Porter untuk Bahasa Indonesia.....	9
2.2.4.    Algoritma K-Medoids.....	13
2.2.5.    Analisis dan Pengujian.....	15
BAB 3.....	16
ANALISIS DAN PERANCANGAN SISTEM.....	16
3.1.    Analisis Data.....	16
3.2.    Data Flow Diagram (DFD).....	18
3.3.    Kamus Data.....	23

## DAFTAR GAMBAR

Gambar 2.1 Tahap-tahap knowledge discovery dalam text mining.....	7
Gambar 2.2 Algoritma Porter untuk Bahasa Indonesia.....	10
Gambar 3.1 Diagram Konteks Sistem Clustering Skripsi Mahasiswa..... dengan Metode K-Medoids.....	18
Gambar 3.2 Data Flow Diagram (DFD) Level 0.....	19
Gambar 3.3 DFD Level 1 – Permintaan Data ke <i>Web Service</i> SInTA.....	20
Gambar 3.4 DFD Level 1 – Text Preprocessing.....	21
Gambar 3.5 Data Flow Diagram (DFD) Level 1 – Perhitungan TF-IDF.....	22
Gambar 3.6 DFD Level 1 – Clustering Dokumen dengan Metode K-Medoids.....	23
Gambar 3.7 Relasi Database Sistem.....	26
Gambar 3.8 Flowchart Proses Kerja Sistem.....	28
Gambar 3.9 Flowchart Proses Permintaan dan Pengolahan Data.....	29
Gambar 3.10 Flowchart Pengambilan Nama Dosen Pembimbing.....	30
Gambar 3.11 Flowchart Tokenisasi.....	31
Gambar 3.12 Filterisasi Stopword.....	32
Gambar 3.13 Proses stemming.....	33
Gambar 3.14 Flowchart Algoritma Porter untuk Bahasa Indonesia.....	34
Gambar 3.15 Flowchart Proses TF-IDF.....	36
Gambar 3.16 Proses Feature Selection.....	37
Gambar 3.17 Flowchart Proses Clustering Skripsi.....	38
Gambar 3.18 Rancangan Halaman Data Skripsi.....	40
Gambar 3.30 Rancangan Halaman Hasil Clustering.....	41
Gambar 4.1 Proses pengambilan data SInTA.....	42
Gambar 4.2 Proses Pengolahan Data SInTA.....	43
Gambar 4.3 Proses Pengambilan Nama Dosen.....	44
Gambar 4.4 Proses Tokenisasi.....	44
Gambar 4.5 Filterisasi Stopword.....	45
Gambar 4.6 Proses Stemming.....	46
Gambar 4.7 Proses Perhitungan TF-IDF.....	47
Gambar 4.8 Proses Normalisasi TF-IDF.....	48
Gambar 4.9 Proses Normalisasi TF-IDF (sambungan).....	49
Gambar 4.10 Proses Feature Selection.....	49
Gambar 4.11 Proses Feature Selection (Sambungan).....	50
Gambar 4.12 Proses Penentuan Centroid.....	51
Gambar 4.13 Proses Clustering.....	52
Gambar 4.14 Proses Clustering(sambungan).....	53
Gambar 4.12 Halaman Proses Clustering SInTA.....	55

## DAFTAR TABEL

Tabel 2.1 Daftar Aturan Partikel.....	11
Tabel 2.2 Daftar Aturan <i>Possessive Pronoun</i> .....	11
Tabel 2.3 Daftar Aturan Awalan Pertama.....	12
Tabel 2.4 Daftar Aturan Awalan Kedua.....	12
Tabel 2.5 Daftar Aturan Akhiran.....	13
Tabel 3.2 Kamus Data.....	24
Tabel 3.2 Kamus Data (Sambungan).....	25
Tabel 4.1 Tabel Pengujian nilai <i>feature selection</i> pada dosen Sistem Informasi.....	58
Tabel 4.2 Tabel Pengujian nilai <i>feature selection</i> pada dosen Teknik Informatika.....	58
Tabel 4.3 Tabel Pengujian jumlah kelompok pada dosen Sistem Informasi.....	62
Tabel 4.4 Tabel Pengujian jumlah kelompok pada dosen Teknik Informatika.....	62



© UKDWN

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

SInTA (Sistem Informasi Tugas Akhir) UKDW merupakan sistem informasi yang menyimpan dan menampilkan dokumen-dokumen skripsi dari mahasiswa UKDW Yogyakarta. Sistem informasi ini memudahkan mahasiswa dalam melakukan pencarian skripsi mahasiswa UKDW Yogyakarta. Skripsi yang ditampilkan dapat dikelompokkan berdasarkan dosen pembimbing, program studi atau bidang minat dari skripsi tersebut. Pengguna dapat melakukan pencarian skripsi dengan memanfaatkan fasilitas pencarian lanjutan. Pencarian dapat dilakukan berdasarkan kata kunci, nama penulis, mata kuliah pendukung, dosen pembimbing, dan bidang minat.

Masalah yang timbul dari pihak dosen adalah ketika dosen ingin mengetahui banyaknya skripsi yang telah dibimbingnya berdasarkan pengelompokan bidang minat skripsi. Dengan mengetahui kelompok dari skripsi yang telah dibimbingnya, maka dosen dapat mengetahui kelompok dengan prosentase terkecil dan mencoba mencari topik skripsi dalam lingkup kelompok tersebut. Untuk pengguna lain, yaitu mahasiswa, tidak sedikit yang mencari referensi skripsi dari SInTA UKDW. Mahasiswa dapat melihat dosen pembimbing dari tiap judul skripsi di SInTA. Masalah timbul ketika mahasiswa akan melakukan konsultasi kepada dosen untuk topik yang skripsi yang akan diajukan. Melalui SInTA, mahasiswa mendapatkan data mengenai dosen pembimbing untuk tiap judul skripsi, sedangkan sebenarnya yang dibutuhkan oleh mahasiswa adalah informasi mengenai dosen yang sering membimbing skripsi dengan topik skripsi yang sejenis dengan milik mahasiswa tersebut.

### **1.2 Perumusan Masalah**

Bagaimana mengelompokkan skripsi mahasiswa menggunakan metode K-Medoids untuk mengetahui topik skripsi bimbingan dosen?

### 1.3 Batasan Masalah

- a. Sistem yang dibangun hanya dapat digunakan untuk pencarian kumpulan skripsi mahasiswa Universitas Kristen Duta Wacana Yogyakarta.
- b. Data yang digunakan dalam penelitian ini adalah data abstrak skripsi mahasiswa program studi Teknik Informatika tahun 2007 sampai dengan 2009 dan program studi Sistem Informasi tahun 2008 sampai dengan 2010.
- c. Parameter yang digunakan dalam penelitian, yaitu dosen pembimbing, abstrak, dan judul skripsi.
- d. Stopword yang digunakan berasal dari <http://lecturer.ukdw.ac.id/budsus/pdf/stopwordID.csv>
- e. Kamus Bahasa Indonesia yang digunakan berasal dari *database* kamus Bahasa Indonesia milik Budi Susanto, S.Kom, M.T.
- f. Metode yang digunakan untuk melakukan perhitungan adalah metode K-Medoids.

### 1.4 Tujuan Penelitian

- a. Membuat program untuk mengelompokkan skripsi mahasiswa.
- b. Mengetahui kecenderungan pengelompokkan topik skripsi mahasiswa.
- c. Melakukan pengujian dan penerapan metode K-Medoids

### 1.5 Metode / Pendekatan

Metode / Pendekatan yang dilakukan dalam penelitian ini, yaitu :

- a. Observasi

Observasi yang dilakukan adalah mengamati dan mempelajari sumber data dari *website* SInTA yang akan digunakan dalam penelitian, terutama melihat kelengkapan dari data yang akan digunakan.

- b. Pengumpulan Data

Pengumpulan data diawali dengan mempelajari cara pengambilan data melalui *web service* SInTA UKDW. Langkah selanjutnya adalah mengakses data yang dibutuhkan melalui *web service* SInTA, kemudian

mencari kamus kata dasar dan kata pelengkap untuk mendukung proses *stemming* dan daftar *stopword*. Langkah akhir adalah melakukan analisis terhadap data SInTA yang telah diperoleh dan mencari kekurangan data yang dibutuhkan.

### c. Pembangunan Sistem

Pembangunan sistem diawali dengan melakukan analisis data dari *web service* SInTA dan menentukan jumlah skripsi yang akan diambil. Selanjutnya menentukan data yang akan diambil, serta mencari solusi untuk kekurangan pada data yang akan diambil. Perancangan sistem dilakukan dengan merancang aliran data pada proses terlebih dahulu, kemudian menentukan data yang akan disimpan dan merancang *database*. Perancangan sistem juga termasuk merancang antarmuka sistem dan merancang alur kerja sistem, seperti proses pengambilan data, *text preprocessing* dan proses *clustering*.

## 1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini secara garis besar dapat dijabarkan menjadi 5 bab. Bab I merupakan bagian pendahuluan yang memberikan gambaran umum mengenai penelitian. Bab ini berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

Tinjauan pustaka yang digunakan dalam penelitian ini dijabarkan pada Bab II. Tinjauan pustaka terdiri dari 2 subbab, yaitu tinjauan pustaka dan landasan teori. Subbab tinjauan pustaka berisi tentang beberapa penelitian yang pernah dilakukan mengenai metode K-Medoids dan penelitian yang memanfaatkan *web service* SInTA. Subbab landasan teori berisi mengenai langkah-langkah yang dilakukan dalam *Text Mining*, pembobotan kata dalam dokumen dengan *Term Frequency-Invers Document Frequency* (TF-IDF), serta algoritma dalam metode K-Medoids berdasarkan pustaka yang digunakan.

Bab III merupakan bagian Analisis dan Perancangan Sistem, yang berisi mengenai penjelasan data yang diperlukan dan digunakan, serta informasi mengenai alat yang akan mendukung penelitian. Perancangan sistem dan desain sistem yang dicantumkan dalam bagian ini, seperti perancangan *database*, flowchart sistem, algoritma yang digunakan, serta rancangan masukan dan keluaran sistem berupa karakteristik masukan yang dibutuhkan dan tampilan keluaran yang dihasilkan.

Deskripsi mengenai Implementasi dan Analisis Sistem dituliskan pada Bab IV. Bagian ini memuat penjelasan mengenai hasil penelitian dan pengujian yang dilakukan, serta melakukan analisis hasil penelitian terhadap tujuan penelitian pada Bab I. Bagian terakhir dari penelitian ini adalah Bab V. Bab V merupakan bagian Kesimpulan dan Saran. Bagian ini akan menjawab rumusan masalah pada Bab I, yang menjadi kesimpulan dari penelitian. Adapun beberapa hal atau pemikiran yang ditemukan pada saat penelitian dan diluar batasan masalah, dapat menjadi saran bagi pengembangan penelitian selanjutnya dengan topik yang sama.





## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan analisis hasil *clustering*, maka didapatkan kesimpulan sebagai berikut:

1. Program yang dibuat mampu mengelompokkan dan mengetahui kecenderungan pengelompokkan skripsi mahasiswa.
2. Metode K-Medoids mengelompokkan skripsi mahasiswa dengan cara menentukan medoid dan *Total Distance* paling kecil dalam kombinasi medoid pada *cluster*, namun pengelompokan ini tergantung pada variasi token yang dihasilkan dari masing-masing dokumen, karena token tersebut merupakan token yang mewakili setiap dokumen dan menentukan jarak antar dokumen.
3. Keragaman jenis dokumen, dan jumlah kelompok dalam *clustering* mempengaruhi nilai *purity*. Nilai *purity* akan semakin tinggi apabila jumlah kelompok yang dipilih semakin banyak dan keragaman jenis dokumen tergolong homogen.

#### 5.2. Saran

Saran yang diberikan untuk pengembangan dan perbaikan sistem agar lebih optimal, yaitu :

1. Pemakaian daftar *stopword* akan lebih maksimal apabila terdapat fasilitas untuk penambahan daftar *stopword*.
2. Proses *stemming* dengan algoritma Porter menghasilkan beberapa kata dasar yang kehilangan maknanya. Hal ini dikarenakan bagian kata dasar dianggap sebagai kata imbuhan. Perbaikan algoritma *stemming* untuk Bahasa Indonesia dengan menggunakan algoritma yang lebih akurat akan menghasilkan kata dasar yang lebih bermakna.
3. Penggunaan kamus kata dasar dan kata pelengkap akan membantu mempercepat proses *stemming*. Akan lebih baik apabila kamus dilengkapi

dengan kata asing sehingga dapat mengurangi kesalahan pada proses stemming yang terdapat dalam teks Bahasa Indonesia.

4. Metode untuk penentuan medoid memiliki pengaruh pada hasil *clustering*, sehingga perbaikan sistem dalam penggunaan metode lain untuk penentuan centroid tanpa menggunakan perkalian matrix akan mempercepat sistem dan memadai untuk nilai kelompok *cluster* yang lebih besar.
5. Jumlah dokumen yang besar akan menghasilkan vektor dokumen dalam jumlah yang besar, sehingga proses text preprocessing akan semakin lama, termasuk pada proses perhitungan bobot token. Perbaikan sistem dalam penggunaan dan pemilihan cara penyimpanan dan pemrosesan data menggunakan teknik lain dapat mempercepat proses.



© UKDW

## DAFTAR PUSTAKA

- Basuki, J.D. (2008). *Implementasi Algoritma K-Medoids untuk Clustering Dokumen Teks*. Program Studi Teknik Informatika Universitas Kristen Duta Wacana. Retrieved from <http://www.sinta.ukdw.ac.id>
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook Advanced Approached in Analyzing Unstructured Data*. New York : Cambridge University Press
- Gunanto, S.G. (2004). *Program Bantu Analisis Data Mining dengan Metode K-Medoid Partitional Clustering Studi Kasus : Hubungan Nilai Tes Penerimaan Mahasiswa Baru dengan Nilai Akademik Mahasiswa Universitas Kristen Duta Wacana*. Program Studi Teknik Informatika Universitas Kristen Duta Wacana. Retrieved from <http://www.sinta.ukdw.ac.id>
- Han, J., & Kamber, M. (2001). *Data Mining Concepts and Techniques*. San Diego : Academic Press.
- Martanti, G.V. (2010). *Implementasi SInTA Web Service dan Amazon Web Service dalam Pencarian Detail Buku Berdasarkan Daftar Pustaka Tugas Akhir*. Program Studi Teknik Informatika Universitas Kristen Duta Wacana. Retrieved from <http://www.sinta.ukdw.ac.id>
- Roberts, D. (May 21, 2009). *K-Medoids: CUDA Implementation*. Access on 17/07/2011 20:30
- Tala, F.Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Netherland : Master of Logic Project Institute for Logic, Language and Computation Universiteit van Amsterdam.3-9
- Weiss, S.M., Indurkha, N., Zhang, T., & Damerau, F.J. (2005). *Text Mining Predictive Methods For Analyzing Unstructured Information*. New York : Springer Science+Bussiness Media, Inc.