



ISSN 2252-9063

*Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika
(KARMAPATI)*

Volume 1, Nomor 4, Agustus 2012

**PENGEMBANGAN APLIKASI *TEXT MINING* DENGAN METODE
ASSOCIATION RULE ANALYSIS UNTUK PENCARIAN DOKUMEN**

Oleh

Kadek Indah Purnama Sari
Jurusan Pendidikan Teknik Informatika, Fakultas Teknik dan Kejuruan,
Universitas Pendidikan Ganesha (Undiksha)
E-mail : indahpunk@yahoo.co.id

ABSTRAK

Penelitian ini bertujuan: (1) Merancang Aplikasi *Text Mining* Dengan Metode *Association Rule* Untuk Pencarian Dokumen.. (2) Mengimplementasikan rancangan Aplikasi *Text Mining* Dengan Metode *Association Rule* Untuk Pencarian Dokumen. Dalam perancangan dan pengimplementasiannya, peneliti menggunakan *Text Mining* dengan Metode *Association Rule Analysis* yang memiliki 3 tahapan: (1) *Parsing Text*, (2) *Stopword*, (3) *Stemming*. Inputan dari aplikasi ini berupa teks berbahasa Indonesia, sedangkan keluarannya hasil pencarian yaitu dokumen terkait. Proses pengujian menggunakan pengujian *blacx box* (fungsional) dan pengujian *white box* (konseptual/struktural).

Hasil dari penelitian ini adalah aplikasi *Text Mining* dengan menggunakan metode *association rule analysis* untuk pencarian dokumen, yang diimplementasikan menggunakan bahasa pemrograman *Delphi 7*. Berdasarkan hasil uji coba, dapat diketahui bahwa “Text Miner” sudah dapat melakukan pencarian berdasarkan *rule* yang didapat dengan baik sehingga dapat ditarik kesimpulan bahwa aplikasi “Text Miner” dapat membantu pencarian dokumen.

Kata kunci: *Text Mining, Association Rule, Stopword, Parsing, Stemming*.



DEVELOPMENT OF TEXT MINING APPLICATION WITH ASSOCIATION RULE METHOD FOR DOCUMENTS SEARCHING

ABSTRACT

This study aimed in: (1) Designing Text Mining Applications With Association Rule Methods for Documents Searching. (2) Implementing Text Mining Applications With Association Rule Methods for Documents Searching. In the design and implementation, researcher was using Text Mining with Association Rule Analysis Method which has three stages: (1) Parsing Text, (2) Stopwording, (3) Stemming. The input of this application is Indonesian language text, whereas the output is searching result which are documents to be searched. The testing process used black box testing (functional) and the white box testing (conceptual / structural).

The results of this research is the Text Mining Applications With Association Rule Methods for Documents Searching named "Text Miner", which is implemented by using the Delphi 7 programming language. Based on trial results, it is known that the application "Text Miner" has been able to do searching well so that it can be concluded that the application "Text Miner" can help to search documents.

Keywords: Text Mining, Association Rule, Stopword, Parsing, Stemming.

1. PENDAHULUAN

Perkembangan teknologi dewasa ini telah mengalami peningkatan yang sangat pesat. Hal ini diiringi juga dengan semakin berkembangnya Teknologi Informasi yang dibutuhkan oleh pengguna sehingga mengakibatkan munculnya suatu cabang ilmu baru dalam teknologi informasi, yaitu pencarian informasi (*information retrieval*). Era informasi telah mengakibatkan usaha penyimpanan data elektronik dalam jumlah yang sangat besar menjadi sulit untuk dilakukan. Perkembangan teknologi yang memungkinkan penemuan, pengembangan, dan pertukaran informasi dalam waktu yang cepat menimbulkan permasalahan yang sama pada repositori data : berlebihnya beban data yang harus disimpan. Hal ini juga menjadi masalah tersendiri bagi pengguna informasi, yaitu sulitnya untuk menemukan informasi yang benar-benar berguna dalam waktu yang singkat dan cara yang nyaman (Marselina, 2009).



Pencarian informasi yang tepat dan sesuai kebutuhan menjadi sangat penting dengan semakin mudahnya memperoleh informasi dari seluruh dunia sebagai akibat perkembangan teknologi informasi dan komunikasi yang semakin pesat, oleh karena itu, teknik untuk memperoleh dokumen dengan isi yang sesuai dengan kebutuhan informasi sangat diperlukan. Sebagai contoh, apabila ada mahasiswa mencari dokumen yang berhubungan dengan *database* dalam artian mahasiswa juga ingin mengetahui informasi lain yang berhubungan dengan *database*, maka seharusnya ditampilkan juga judul dokumen yang dimana isi abstraknya mengandung kata penting *database*, beserta dokumen lainnya yang mengandung kata penting yang berhubungan atau berasosiasi dengan kata penting *database*, misalnya *Table*, *DFD*, *ERD*, *Query*, dll (Ferry,2005)

Untuk membantu memecahkan masalah pencarian tersebut, tentunya diperlukan sebuah solusi, dimana solusi tersebut nantinya diharapkan bisa membantu *user* dalam melakukan pencarian informasi yang bermakna. *Text Mining* adalah salah satu solusi yang dapat membantu permasalahan diatas. *Text Mining* merupakan bagian dari *Data Mining*, dimana yang digali disini adalah data yang berupa teks yang biasanya ditemukan dalam dokumen. *Text Mining* merupakan bidang baru yang mencoba untuk mengumpulkan informasi bermakna dari teks bahasa alami. *Text Mining* identik sebagai proses menganalisis teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu. Dibandingkan dengan jenis data yang disimpan dalam database, teks tidak terstruktur, *amorphous*, dan sulit untuk ditangani menggunakan algoritma. Namun, dalam budaya modern, teks adalah kendaraan umum untuk pertukaran informasi formal. Bidang *Text Mining* biasanya berhubungan dengan teks yang berfungsi sebagai komunikasi informasi faktual atau pendapat (Ian, 2000).

Dalam penelitian ini, peneliti akan membuat sebuah Aplikasi *Text Mining* dengan menggunakan metode *Association Rule Analysis*. Peneliti memilih topik ini karena berdasarkan jurnal yang penulis dapatkan mengenai *Text Mining*, pencarian yang dilakukan masih berpatokan pada suatu kategori tertentu, beda halnya dengan



aplikasi yang akan peneliti buat, aplikasi ini akan melakukan pencarian secara global berdasarkan masukan *keyword* yang diberikan oleh *user*. Aplikasi ini nantinya akan melakukan penyaringan kata-kata tidak penting, penyimpanan kata penting pada tiap abstrak, menemukan dan menganalisa relasi antar abstrak, dan kemudian menampilkan hasil analisa tersebut kepada *user*. Dengan aplikasi ini nantinya diperoleh daftar judul dokumen dari yang paling berhubungan dengan *keyword* yang diinputkan oleh user sampai dengan yang paling sedikit hubungannya dengan *keyword* yang diinputkan oleh user.

2. METODE

2.1 Text Mining

Kebanyakan fokus penelitian *Data Mining* sebelumnya terletak pada data-data yang sifatnya terstruktur, seperti relasional, transaksi dan data-data *warehouse*. Namun, pada kenyataannya, sebagian besar informasi yang tersedia, tersedia dalam bentuk teks *database* (*database* dokumen) yang terdiri dari koleksi besar dokumen dari berbagai sumber, seperti artikel berita, makalah penelitian, buku, perpustakaan digital, e-mail, dan halaman web (Han, Kamber,2006).

Sama halnya seperti *Data Mining* yang digunakan untuk mencari pola dari suatu data, *Text Mining* juga digunakan untuk mencari pola data yang berupa teks, walaupun keduanya terlihat sama, namun terdapat perbedaan nyata diantara keduanya. *Data Mining* lebih identik dengan proses ekstraksi dari data yang implicit, informasi yang sebelumnya tidak diketahui, sedangkan pada *Text Mining*, informasi yang akan diekstrak adalah jelas dan eksplisit dinyatakan dalam teks tersebut (Ian, 2000). Dalam *Text Mining* terdapat beberapa tahapan-tahapan penting seperti misalnya, *Parsing*, *Filtering*, dan *Stemming*.



2.2 Association Rule Analysis

Association Rule adalah metode untuk menemukan hubungan di antara data. Meskipun tidak semua *rule* yang dihasilkan mencerminkan hubungan yang sebenarnya di dunia nyata, metode ini dapat membantu mengenali pola-pola tertentu di dalam kumpulan data yang besar, misalnya mengetahui pola pembelian barang di suatu supermarket dimana suatu barang dibeli setiap kali suatu barang yang lain dibeli juga.

Sebagai contoh, *customer* yang biasanya membeli komputer juga membeli *financial_management_software* ditunjukkan sebagai berikut :

Computer -> *financial_management_software* [*support* = 2%, *confidence* = 60%]

Dalam pencarian *association rule*, diperlukan suatu variable ukuran yang dapat ditentukan oleh *user*, untuk mengatur batasan sejauh mana dan sebanyak apa hasil output yang diinginkan *user*. *Support* dan *confidence* adalah dua ukuran kepercayaan yang menunjukkan kepastian dan tingkat kegunaan suatu *rule* yang ditemukan. *Support* 2% berarti bahwa sebanyak 2% dari seluruh transaksi yang dianalisis menunjukkan bahwa komputer dan *financial_management_software* dibeli secara bersamaan. Sedangkan *confidence*=60%, berarti bahwa 60% dari *customers* yang membeli sebuah computer pasti juga membeli *software*.

3. ANALISIS DAN PERANCANGAN

3.1 Analisis Masalah dan Usulan Solusi

Masalah yang akan diteliti dalam perangkat lunak ini adalah proses mencari dokumen berdasarkan atas kata kunci masukan dari user dengan menggunakan konsep *text mining* beserta metode dalam *data mining* yaitu metode *association rule analysis*. *Text mining* sendiri adalah sama dengan data mining, hanya saja dalam *text mining* yang digali adalah data yang berupa *text* yang biasanya banyak terdapat dalam dokumen. Pada umumnya, fasilitas *search engine* yang peneliti temukan cenderung



masih menggunakan metode lama artinya pencariannya masih berpatokan pada kategori tertentu misalnya judul buku, penulis, dan lain-lain, sehingga informasi dari hasil pencarian ini hanya akan memberikan sedikit informasi kepada *user*, sebagai contoh, apabila ada mahasiswa mencari dokumen yang berhubungan dengan *database* dalam artian mahasiswa juga ingin mengetahui informasi lain yang berhubungan dengan *database*, maka seharusnya ditampilkan juga judul dokumen yang dimana isi abstraknya mengandung kata penting *database*, beserta tugas akhir lainnya yang mengandung kata penting yang berhubungan atau berasosiasi dengan kata penting *database*, misalnya *Table, DFD, ERD, Query*, dll (Ferry,2005)

Untuk membantu memecahkan masalah pencarian tersebut, tentunya diperlukan sebuah solusi, dimana solusi tersebut nantinya diharapkan bisa membantu *user* dalam melakukan pencarian informasi yang bermakna. *Text Mining* adalah salah satu solusi yang dapat membantu permasalahan diatas, dalam aplikasi ini nantinya ada dua proses penting yang menjadi inti dari perangkat lunak ini, yaitu proses *preprocessing* abstrak dokumen dan proses *generate association rule* yang bertujuan untuk menemukan aturan-aturan yang nantinya akan digunakan pada proses pencarian dokumen oleh *user*.

3.2 Analisis Perangkat Lunak

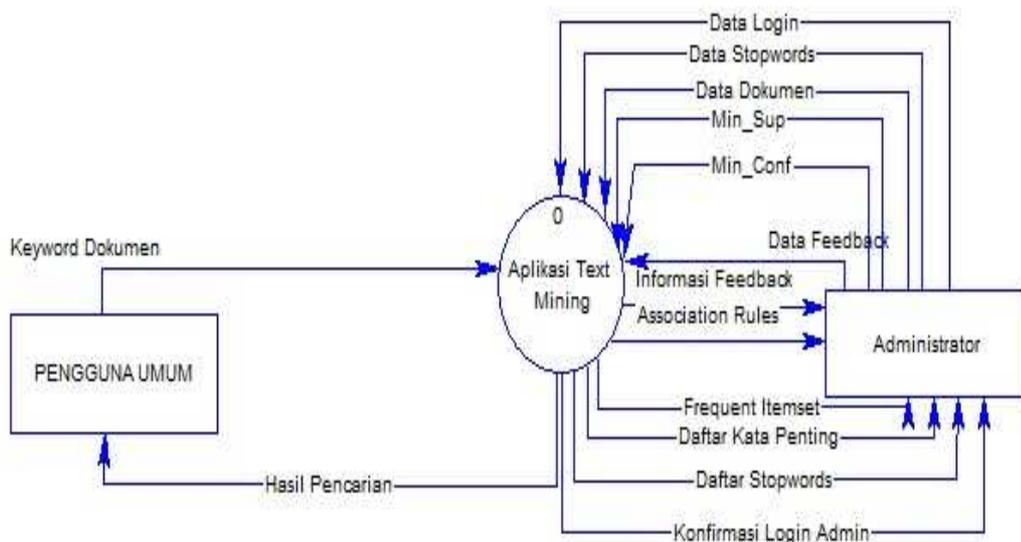
Secara umum perangkat lunak aplikasi *Text Mining* ini diharapkan memiliki beberapa fungsi utama yaitu menginputkan data dokumen, melakukan *preprocessing* seperti melakukan proses *parsing, stopword removal, stemming* dan *feedback*, kemudian proses meng-*generate frequent itemset, generate rule* dan melakukan proses pencarian dokumen.

Tujuan dari pengembangan perangkat lunak ini adalah dapat menginputkan data dokumen, melakukan *preprocessing* seperti melakukan proses *parsing, stopword removal, stemming* dan *feedback*, kemudian proses meng-*generate frequent itemset, generate rule* dan melakukan proses pencarian dokumen.

Masukan dari aplikasi ini adalah dokumen berformat *.txt, *.doc, *.docx, *.rtf, data admin, data dokumen, nilai *min_sup_count*, nilai *minimum support*, nilai *minimum confidence*, *user's keyword*, data *feedback*, data *stopwords*. Sedangkan keluaran dari sistem ini adalah Informasi *Login*, informasi *stopwords*, informasi dokumen, hasil pencarian dokumen, informasi aturan asosiasi, informasi *feedback*, informasi tentang program, dan informasi tentang pembuat program.

3.3 Perancangan Perangkat Lunak

Untuk menggambarkan alur perpindahan data dalam sistem digunakan DFD. DFD atau *Data Flow Diagram* adalah diagram aliran data yang sering digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem baru yang akan dikembangkan secara logika tanpa pertimbangan lingkungan fisik dimana data tersebut mengalir (misalnya lewat telepon, surat dan sebagainya) atau lingkungan fisik dimana data tersebut disimpan. Berikut ini akan digambarkan Diagram Konteks atau DFD Level 0 dan arsitektur perangkat lunak yang akan dibangun.

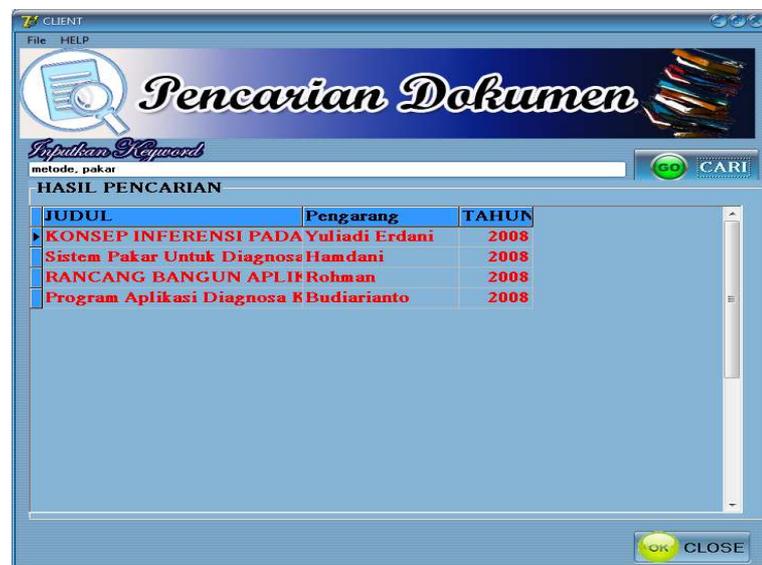


Gambar 1. Diagram Konteks Aplikasi *Text Mining*

4. IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Perangkat Lunak

Data Flow Diagram (DFD) dan Rancangan Arsitektur Perangkat Lunak Aplikasi *Text Mining* diimplementasikan dengan menggunakan bahasa pemrograman *Delphi7*. Tampilan *form* utama Aplikasi *Text Miner* dapat dilihat pada gambar 2.



Gambar 2. Tampilan Utama Aplikasi *Text Miner* untuk Pencarian Dokumen

4.2 Pengujian Perangkat Lunak

Tabel 1. Pelaksanaan Uji Kebenaran hasil *stopword removal* pada aplikasi *Text Miner*

No	Abstrak Dokumen Uji	Hasil <i>Stopword Removal</i>
1.	Kunci untuk meningkatkan kepuasan pelanggan adalah mempertahankan serta meningkatkan pelayanan kemudian memberikan pelayanan yang baik pada pelanggan (Help-Desk (HD)). Salah satu cara meningkatkan mutu pelayanan kepada pelanggan yaitu dengan menyediakan bantuan melalui jalur telepon	kunci meningkatkan kepuasan pelanggan mempertahankan meningkatkan pelayanan pelanggan help desk hd salah meningkatkan mutu pelayanan pelanggan menyediakan bantuan jalur telepon ditangani help desk

No	Abstrak Dokumen Uji	Hasil <i>Stopword Removal</i>
	<p>yang ditangani oleh help desk operators. Atas dasar peningkatan kepuasan pelanggan maka sebuah sistem Open source yang mampu mendiagnosa kerusakan mobil dengan menalarakan solusi berbasis kasus (CBR) terdiri dari: a case base, a retriever, an adapter, a refiner, an executer dapat diterapkan untuk membantu mempercepat proses pelayanan terhadap pelanggan, yang dapat diaplikasikan oleh customer service / help desk Officer.</p>	<p>operators dasar peningkatan kepuasan pelanggan sistem open source mendiagnosa kerusakan mobil menalarakan solusi berbasis cbr a case base a retriever an adapter a refiner an executer diterapkan mempercepat proses pelayanan pelanggan diaplikasikan customer service help desk officer</p>

Tabel 2. Pelaksanaan Uji Kebenaran hasil *stemming* pada aplikasi Text Miner

No	Abstrak Dokumen Uji	Hasil <i>Stemming</i>	<i>Feedback</i>	Analisis
1.	<p>Penyakit THT(Telinga Hidung Tenggorokan) merupakan penyakit yang sering terjadi dimasyarakat. Sistem pakar merupakan salah satu solusi untuk mendiagnosis penyakit berdasarkan gejala yang dirasakan oleh penderita.</p>	<p>sakit tht telinga hidung tenggoro sakit masyarakat sistem pakar salah solusi diagnosis sakit dasar gejala rasa derita teliti sistem pakar konsep forward chaining metode certainty factor cf faktor pasi diagnosa sakit tht manusia sistem diagnosa sakit tht derita derit gejala gejala rasa derita langsung pakar hasil teliti cf atas tidakpasti diagnosa tht</p>	<p>sakit tht telinga hidung tenggorokan sakit masyarakat sistem pakar salah solusi diagnosis sakit dasar gejala rasa derita teliti sistem pakar konsep forward chaining metode certainty factor cf faktor pasi diagnosa sakit tht manusia sistem diagnosa</p>	<p>Hasil <i>stemming</i> sesuai konteks</p>

No	Abstrak Dokumen Uji	Hasil Stemming	Feedback	Analisis
			sakit tht derita derit gejala gejala rasa derita langsung pakar hasil teliti cf atas tidakpasti diagnosa tht	

Tabel 3. Pelaksanaan Uji Kebenaran hasil *searching* pada aplikasi *Text Mining*

No.	Keyword Input	Hasil Pencarian (Kode Dokumen)	Analisis
1.	Data Mining	D2	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D2, karena pada abstrak dokumen D2 mengandung kata “ data ” dan “ mining ”.
		D17	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D17, karena pada abstrak dokumen D17 mengandung kata “ data ” dan “ mining ”.
		D1	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D1, karena pada abstrak dokumen D1 mengandung kata “ data ”.
		D4	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D4, karena pada abstrak

No.	Keyword Input	Hasil Pencarian (Kode Dokumen)	Analisis
			dokumen D4 mengandung kata “data” .
		D5	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D5, karena pada abstrak dokumen D5 mengandung kata “data” .
		D13	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D13, karena pada abstrak dokumen D13 mengandung kata “data” .
		D15	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D15, karena pada abstrak dokumen D15 mengandung kata “data” .
		D16	Pencarian dokumen dengan keyword data mining ”, menghasilkan dokumen dengan kode dokumen D16, karena pada abstrak dokumen D16 mengandung kata “data” .

Tabel 4. Pelaksanaan Uji Algoritma pada Aplikasi *Text Mining*

No.	Nama Algoritma	Tanggal Pengujian			
		7 Juni 2012		30 Juni 2012	
		Sesuai	Tidak Sesuai	Sesuai	Tidak Sesuai
1.	<i>Parsing Abstrak</i>	√		√	
2.	<i>Stopword Removal</i>	√		√	
3.	<i>Stemming</i>	√		√	
4.	Generate Frequent Itemset	√		√	

Berdasarkan hasil pengujian fungsional menunjukkan bahwa sistem sudah bisa menangani data input yang tidak valid dan menampilkan output sesuai dengan apa yang direncanakan. Hasil pengujian konseptual menunjukkan bahwa sistem telah



melaksanakan mekanisme perhitungan dan mekanisme logika sesuai dengan apa yang direncanakan.

5. PENUTUP

5.1 Simpulan

Simpulan yang dapat diambil dari pembahasan yang telah dikemukakan pada bab sebelumnya adalah sebagai berikut :

1. Implementasi *Text Mining* Menggunakan Metode *Association Rule* untuk Pencarian Dokumen menghasilkan suatu program aplikasi yang menangani proses-proses berikut ini :
 - a. Manipulasi Data
Sistem menangani proses manipulasi data dokumen, data *stopword*, dan data admin.
 - b. Pencarian data (Dokumen)
Sistem menangani proses pencarian data dalam hal ini dokumen berdasarkan *keyword* atau kata kunci yang diinputkan oleh user, pencarian dilakukan dengan mencocokkan kata kunci user dengan database sistem.
2. Dari hasil pengujian yang meliputi pengujian fungsional dan pengujian konsep/struktural, dapat disimpulkan bahwa sistem bisa menangani data input yang tidak valid dan menampilkan output sesuai dengan apa yang direncanakan serta melaksanakan mekanisme perhitungan dan mekanisme logika sesuai dengan apa yang direncanakan.

5.2 Saran

Penulis berharap agar Aplikasi *Text Mining* ini bisa disempurnakan dan dikembangkan lebih jauh sehingga menjadi lebih lengkap dan sempurna. Adapun saran yang dapat diberikan dalam upaya penyempurnaan dan pengembangan sistem adalah sebagai berikut :



1. Diharapkan validasi saat admin melakukan login bisa ditingkatkan lagi keamanannya, sehingga otoritas admin dapat terjamin dan data yang ditangani oleh admin bersangkutan tidak dapat diubah oleh orang lain.
2. Diharapkan agar nantinya sistem mampu dikembangkan berbasis web.

6. DAFTAR PUSTAKA

- Han, Jiawei, and Micheline Kamber. Data Mining : Concepts and Techniques. Morgan Kaufmann, 2006.
- Marlina, Euis. 2009. 10 Jenis Koneksi Delphi ke Database. Cetakan Ke-1. Yogyakarta : Gava Media
- Tala, Fadillah Z. A Study of Stemming Effects on Information Retrieval In Bahasa Indonesia. Institute for Logic, Language and Computation Universeit Van Amsterdam, 2003
- Tando, Marselina. 2009. Teknologi Text-Mining untuk Pengembangan Ilmu Biologi sebagai Ilmu Interdisipliner di Masa Depan. Terdapat pada : <http://jurnal.ui.ac.id>. Diakses tanggal 15 Pebruari 2012.
- Sucahyo, Yudo Giri. 2003. "Data Mining : Menggali Informasi Yang Terpendam". Terdapat pada : <http://ilmukomputer.com>. Diakses tanggal 15 Pebruari 2012.
- Susanto, Sani dan Dedy Suryadi, S.T., MS. 2010. Pengantar Data Mining . Cetakan Ke-1. Yogyakarta : Andi.
- Witten, Ian H. 2000."Text Mining" . Terdapat pada : <http://www.cs.waikato.ac.nz/~nzdl/textmining>. diakses tanggal 4 januari 2012.
- Yuwono, Ferry et.al. 2005. "Algoritma Porter Stemmer for Bahasa Indonesia untuk Pre-processing Text Mining Berbasis Metode Market Basket Analysis". Terdapat pada : <http://www.fportfolio.petra.ac.id>. Diakses tanggal 10 Januari 2012.