

**ANALISIS ASOSIASI KUMPULAN KATA KUNCI PADA  
ABSTRAKSI DOKUMEN KARYA TULIS ILMIAH STUDI  
KASUS : SINTA UKDW.**

Tugas Akhir



oleh  
**JOHANES SITUMORANG**  
**22074316**

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

**ANALISIS ASOSIASI KUMPULAN KATA KUNCI PADA  
ABSTRAKSI DOKUMEN KARYA TULIS ILMIAH STUDI  
KASUS : SINTA UKDW.**

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

**JOHANES SITUMORANG**  
**22074316**

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

## PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul:

ANALISIS ASOSIASI KUMPULAN KATA KUNCI PADA ABSTRAKSI  
DOKUMEN KARYA TULIS ILMIAH STUDI KASUS: SINTA UKDW

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 yang sumber informasinya dicantumkan sebagaimana mestinya.

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

Yogyakarta,



(Johanes Situmorang)

22074316

## HALAMAN PENGESAHAN

### ANALISIS ASOSIASI KUMPULAN KATA KUNCI PADA ABSTRAKSI DOKUMEN KARYA TULIS ILMIAH STUDI KASUS : SINTA UKDW.

Oleh: JOHANES SITUMORANG / 22074316

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

Yogyakarta, 20 Juni 2012  
Mengesahkan,

Dewan Penguji:

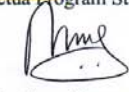
1. Budi Susanto, SKom., M.T.
2. Rosa Delima, S.Kom., M.Kom.
3. Kathryn Widhiyanti, M.Cs.
4. Aloysius Airlangga Banuadji, S.Kom., M.Eng.

**DUTA WACANA**

Dekan

Ketua Program Studi

  
(Drs. Wimmie Handiwidjojo, M.T.)

  
(Nugroho Agus Haryono, M.Si)

## UCAPAN TERIMA KASIH

Puji syukur dan terima kasih yang utama penulis naikan kepada Tuhan Yesus Kristus yang selalu mengasihi serta memberi kekuatan dalam keadaan susah maupun senang, yang selalu memberikan petunjuk dalam setiap kegalauan selama menyelesaikan tugas akhir ini sehingga tugas akhir ini dapat selesai. Ucapan terima kasih juga penulis naikan kepada kedua orang tua penulis yang selalu memberi kekuatan, doa serta penghiburan dalam proses pengerjaan tugas akhir ini semoga Tuhan Yesus selalu memberi kesehatan kepada beliau. Selanjutnya ucapan terima kasih juga ditujukan kepada Budi Susanto, S.Kom.,M.T selaku dosen pembimbing I dan Rosa Delima, M.Kom selaku dosen pembimbing II yang sudah membimbing, memberi pengarahan dan tambahan ilmu dalam penyelesaian tugas akhir ini semoga Tuhan Yesus memberkati. Ucapan terima kasih juga penulis sampaikan kepada rekan-rekan *crew* unit PPTPM UKDW yang selalu memberi motivasi dan dukungan untuk tetap semangat dalam menyelesaikan tugas akhir ini serta kepada pihak-pihak yang tidak dapat penulis sebutkan satu persatu.

Akhir kata penulis berharap semoga laporan ini dapat berguna dan dapat memberikan masukan bagi para pembaca.

Yogyakarta, 28 Mei 2012

Penulis

## ABSTRAKSI

Dokumen teks pada umumnya merupakan suatu dokumen yang terdiri dari rangkaian kata-kata yang terbentuk menjadi suatu informasi atau pengetahuan. Dari rangkaian kata-kata tersebut biasanya akan terdapat beberapa kata yang mempunyai arti khusus. Kata yang dimaksud itu adalah kata kunci. Dari kumpulan kata kunci itu akan ada yang sering muncul bersamaan di dalam setiap dokumen. Kemunculan kata kunci yang sering bersamaan itu dapat menjadi suatu pola asosiatif yang bermanfaat. Untuk mengetahui seberapa kuatnya pola-pola asosiatif kata kunci yang sering muncul bersamaan, dapat diukur dengan teknik dalam *Data mining* yang disebut *Association rule*.

Pembentukan pola-pola asosiatif dengan *association rule* menggunakan algoritma *fp-growth* dalam pencarian *frequent itemset* nya dengan cara membentuk pohon data yang berupa kata kunci hasil dari *text preprocessing* abstraksi suatu dokumen karya ilmiah mahasiswa UKDW yang digunakan sebagai data sampel.

Hasil analisis dari pembentukan pola asosiasi kata kunci dokumen abstraksi ini menghasilkan pola-pola kata kunci yang akurat pada data sampel namun ketika diujikan ke data uji banyak yang kurang akurat. Hal ini disebabkan data abstraksi suatu dokumen karya ilmiah banyak menggunakan kata kunci-kata kunci yang unik yang muncul cenderung hanya di dalam abstraksi dirinya sendiri.

**Kata kunci :** *Data Mining, Text Mining, Association Rule, FP-Growth, Chi-Square*

## DAFTAR ISI

HALAMAN JUDUL .....	
PERNYATAAN KEASLIAN TUGAS AKHIR .....	i
HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
UCAPAN TERIMA KASIH .....	iv
ABSTRAKSI.....	v
DAFTAR ISI .....	vi
DAFTAR TABEL .....	ix
DAFTAR GAMBAR .....	x
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Pendahuluan .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	2
1.5 Metode Penelitian .....	3
1.6 Sistematika Penulisan .....	3
<b>BAB 2 TINJAUAN PUSTAKA .....</b>	<b>5</b>
2.1 Tinjauan Pustaka .....	5
2.2 Landasan Teori .....	6
2.2.1 Data Mining .....	6
2.2.2 Text Mining .....	7
2.2.3 Association Rule Mining .....	7
2.2.4 Frequent Pattern Growth .....	9
2.2.5 Uji Chi-Square .....	14
<b>BAB 3 ANALISIS DAN PERANCANGAN SISTEM .....</b>	<b>16</b>
3.1 Analisis Data .....	16

3.2 Use Case Diagram .....	17
3.3 Kamus Data .....	18
3.4 Alur Kerja Sistem .....	24
3.4.1 Permintaan Data Skripsi ke Web Service SInTA .....	24
3.4.2 Penyaringan Data Abstraksi .....	25
3.4.3 Pembuatan Abstraksi Sampel .....	26
3.4.4 Pembuatan Pasangan Kata .....	27
3.4.5 Pembentukan Kata Kunci .....	28
3.4.6 Pencarian Pola Asosiasi Kata Kunci .....	29
3.5 Rancangan Hasil Sistem .....	31
3.5.1 Tampilan Untuk Keperluan Analisis .....	31
3.5.2 Tampilan Untuk Pengguna .....	31
<b>BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM .....</b>	<b>33</b>
4.1 Implementasi Sistem .....	33
4.1.1 Proses Pengambilan Data SInTA .....	33
4.1.2 Proses Pembuatan Data Sampel .....	33
4.1.3 Proses Pembuatan Pasangan Kata .....	35
4.1.4 Proses Pembuatan Kata Kunci .....	36
4.1.5 Proses Pembentukan Frequent Itemset Kata Kunci .....	40
4.1.6 Proses Pembentukan Pola Asosiasi Kata Kunci .....	44
4.2 Hasil Implementasi .....	44
4.2.1 Pembentukan Kata Kunci .....	44
4.2.2 Pembentukan FP-Tree.....	47
4.2.3 Pembentukan Single Path(Lintasan Tunggal) .....	49
4.2.4 Pembentukan Frequent Itemset .....	49
4.2.5 Hasil Pembentukan Pola Asosiasi .....	50
4.2.6 Proses Pencarian Pola Asosiasi Kata Kunci .....	51



4.3 Analisis Sistem .....	53
4.4 Hasil Analisis .....	61
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>	<b>63</b>
5.1 Kesimpulan .....	63
5.2 Saran .....	63
Daftar Pustaka .....	64

## LAMPIRAN

### A. Source Code Program

©UKDW

## DAFTAR TABEL

Tabel 2.1	Model Transaksi Data .....	8
Tabel 2.2	Tabel Transaksi T .....	10
Tabel 2.3	Frekuensi Kemunculan Tiap karakter .....	10
Tabel 2.4	Kemunculan Tiap Item dalam Setiap Transaksi .....	11
Tabel 2.5	Hasil Pembentukan Conditional fp-tree .....	14
Tabel 2.6	Tabel Kontingensi .....	15
Tabel 3.1	Kamus Data .....	18
Tabel 4.1	Tabel Observasi Pasangan Kata “Kecerdasan Buatan”... ..	37
Tabel 4.2	Tabel Observasi Pasangan Kata “Operasi Linux” .....	37
Tabel 4.3	Tabel Harapan Pasangan Kata “Kecerdasan Buatan” ....	38
Tabel 4.4	Tabel Harapan Pasangan kata “Operasi Linux” .....	39
Tabel 4.5	Dua Puluh Kata Kunci yang Mempunyai Frekuensi Terbesar .....	45
Tabel 4.6	Sebaran Frekuensi Kata Kunci .....	46
Tabel 4.7	Hasil Pengujian Pola Asosiasi Kata Kunci “kecerdasan buatan” dengan Minimum <i>Support</i> 2 .....	53
Tabel 4.8	Hasil Pengujian Pola Asosiasi Kata Kunci “perangkat lunak” dengan Minimum <i>Support</i> 3 .....	57
Tabel 4.9	Hasil Pengujian Pola Asosiasi Kata Kunci “borland delphi” dengan Minimum <i>Support</i> 4 .....	58
Tabel 4.10	Hasil Pengujian Pola Asosiasi Kata Kunci “sistem informasi” dengan Minimum <i>Support</i> 5 .....	58

## DAFTAR GAMBAR

Gambar 2.1	FP-Tree .....	11
Gambar 2.2	Pembentukan <i>fp-tree</i> untuk TID 01 .....	12
Gambar 2.3	Pembentukan <i>fp-tree</i> untuk TID 02 .....	12
Gambar 2.4	Pembentukan <i>fp-tree</i> untuk TID 05 .....	13
Gambar 3.1	<i>Use Case Diagram</i> .....	17
Gambar 3.2	Relasi <i>Database</i> Sistem .....	23
Gambar 3.3	Diagram Alur Proses Kerja Sistem Secara Umum .....	24
Gambar 3.4	Diagram Alur Proses Permintaan Data Skripsi ke <i>Web Service</i> SInTA .....	25
Gambar 3.5	Diagram Alur Proses Penyaringan Data Skripsi .....	26
Gambar 3.6	Diagram Alur Proses Pembentukan abstraksi sampel .....	27
Gambar 3.7	Diagram Alur Proses Pembuatan Pasangan Kata .....	28
Gambar 3.8	Diagram Alur Proses Pembentukan Kata Kunci .....	29
Gambar 3.9	Diagram Alur Proses Pencarian Pola Asosiasi Kata Kunci .....	30
Gambar 3.10	Rancangan Tampilan Hasil untuk Keperluan Analisis .....	31
Gambar 3.11	Rancangan Tampilan untuk Pengguna Memasukkan Nilai Minimum <i>Support</i> dan Minimum <i>Confidence</i> .....	32
Gambar 3.12	Rancangan Tampilan untuk Pengguna Memilih Kata Kunci .....	32
Gambar 3.13	Rancangan Tampilan Hasil Pola Asosiasi Kata Kunci untuk Pengguna .....	32
Gambar 4.1	Proses Pengambilan Data SInTA .....	33
Gambar 4.2	Proses Pembuatan Data Sampel .....	34
Gambar 4.3	Proses Pembuatan Pasangan Kata .....	35
Gambar 4.4	Proses Perhitungan Nilai Observasi Pasangan Kata .....	36
Gambar 4.5	Proses Perhitungan Nilai Harapan Pasangan Kata .....	38

Gambar 4.6	Proses Pembentukkan Data Transaksi .....	41
Gambar 4.7	Proses Pembentukkan Pohon Data .....	42
Gambar 4.8	Proses Pemampatan <i>Single Path</i> .....	43
Gambar 4.9	Proses Pembentukan Pola Asosiasi Kata Kunci .....	44
Gambar 4.10	Hasil Pembentukan <i>FP-Tree</i> dengan Minimum <i>Support</i> 2 (kiri) dan Minimum <i>Support</i> 3 (kanan) .....	47
Gambar 4.11	Hasil Pembentukan <i>FP-Tree</i> dengan Minimum <i>Support</i> 4 (kiri) dan Minimum <i>Support</i> 5 (kanan) .....	48
Gambar 4.12	Hasil Pembentukan <i>Single Path</i> .....	49
Gambar 4.13	Hasil Pembentukan <i>Frequent Itemset</i> .....	50
Gambar 4.14	Hasil Pembentukan Pola Asosiasi .....	50
Gambar 4.15	Tampilan <i>Form</i> Minimum <i>support</i> dan Minimum <i>Confidence</i> .....	51
Gambar 4.16	Tampilan <i>Form</i> Pilih Kata Kunci .....	52
Gambar 4.17	Tampilan Halaman Hasil Pencarian Pola Asosiasi Kata Kunci .....	52
Gambar 4.18	Potongan Abstraksi Data Uji yang Memuat Pola Asosiasi Kata Kunci “ <i>if</i> sistem kecerdasan <i>then</i> kecerdasan buatan” .....	59
Gambar 4.19	Potongan Abstraksi Data Uji yang Memuat Pola Asosiasi Kata Kunci “ <i>if</i> borland delphi^kecerdasan buatan <i>then</i> bahasa pemrograman” .....	59
Gambar 4.20	Potongan Abstraksi Data Uji yang Memuat Pola Asosiasi Kata Kunci “ <i>if</i> kecerdasan buatan^perangkat lunak <i>then</i> bahasa pemrograman” Namun Hanya Premis <i>if</i> nya Saja .....	60

# BAB 1

## PENDAHULUAN

### 1.1 Pendahuluan

Dokumen teks pada umumnya merupakan suatu dokumen yang terdiri dari rangkaian kata-kata yang terbentuk menjadi suatu informasi atau pengetahuan. Dari rangkaian kata-kata tersebut biasanya akan terdapat beberapa kata yang mempunyai arti khusus. Kata yang dimaksud itu adalah kata kunci. Dari kumpulan kata kunci itu akan ada yang sering muncul bersamaan di dalam setiap dokumen. Kemunculan kata kunci yang sering bersamaan itu dapat menjadi suatu pola asosiatif yang bermanfaat. Untuk mengetahui seberapa kuatnya pola-pola asosiatif kata kunci yang sering muncul bersamaan, dapat diukur dengan teknik dalam *Data mining* yang disebut *Association rule*.

*Association Rule* merupakan teknik untuk mengukur seberapa kuat keterikatan dua buah item atau lebih yang terdapat di dalam suatu transaksi data. Misalnya dalam kasus di swalayan, *association rule* mengukur seberapa kuat asosiasi/hubungan, jika pembeli membeli roti bersamaan dengan mentega maka akan membeli susu juga. Hasil analisis dari kasus swalayan tadi berguna bagi pemilik swalayan dalam peletakan item barangnya agar meletakkan barang-barang yang sering dibeli pengunjung bersamaan, selanjutnya akan di letakan berdekatan.

Dalam penelitian ini, penulis mencoba menggunakan *association rule* dalam menganalisis dan menemukan keterikatan antara kata kunci dalam dokumen teks. Dengan mengetahui kumpulan kata kunci yang sering terjadi bersamaan, maka akan dengan mudah bagi pembaca untuk mencari topik lain yang masih berhubungan dengan dokumen yang sedang dibaca.

Dokumen yang dipilih penulis adalah dokumen abstraksi skripsi mahasiswa UKDW (Universitas Krsiten Duta Wacana) yang terdiri dari dua program studi yang berbeda, yaitu Teknik Informatika dan Sistem Informasi.

## 1.2 Rumusan Masalah

Rumusan masalah penelitian ini adalah:

Bagaimana menganalisis asosiasi kumpulan kata kunci pada abstraksi dokumen teks karya ilmiah yang terdapat di Sinta UKDW dengan menerapkan teknik *association rule*?

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Sistem hanya menganalisis kata kunci yang terdapat di dalam abstraksi dokumen karya ilmiah UKDW dari program studi Teknik Informatika dan Sistem Informasi yang bersumber dari situs [sinta.ukdw.ac.id](http://sinta.ukdw.ac.id).
- b. Kumpulan kata kunci yang digunakan sebagai proses analisis terdiri dari 500 buah dokumen yang diambil secara acak dari situs [sinta.ukdw.ac.id](http://sinta.ukdw.ac.id).
- c. Hasil akhir dari sistem hanya akan menampilkan pola asosiasi kumpulan kata kunci yang sering muncul bersamaan pada setiap dokumen yang dianalisis.
- d. Setiap kata kunci yang dianalisis adalah kata kunci yang hanya terdiri dari dua kata.
- e. Pembuatan kata kunci berdasarkan pengujian *chi-square*

## 1.4 Tujuan Penulisan

Dengan menganalisis asosiasi kumpulan kata kunci yang terdapat di dalam abstraksi dokumen karya ilmiah UKDW, maka akan ditemukan pola-pola kata kunci yang sering muncul bersamaan pada setiap abstraksi sehingga dapat digunakan dalam pengembangan SInTA untuk fitur pencarian saran topik-topik yang berhubungan.

## 1.5 Metodologi Penelitian

Metode yang digunakan untuk menyelesaikan penelitian ini adalah:

- a. Studi pustaka, studi pustaka ini diperlukan untuk menunjang pengumpulan informasi mengenai pokok bahasan yang berkaitan langsung dengan *data mining*, *text mining*, *association rule*, algoritma *fp-growth* dan algoritma *chi-square* yang diambil dari buku dan jurnal baik yang terdapat di perpustakaan UKDW ataupun dari media elektronik seperti internet.
- b. Konsultasi dengan dosen pembimbing, melakukan konsultasi dengan dosen pembimbing berkaitan dengan tahap perancangan, pembuatan sistem dan pelaporan hasil penelitian.
- c. Observasi, mengamati dan mempelajari sumber data yang akan diambil untuk keperluan analisis dan pembangunan sistem.
- d. Pembangunan sistem yang meliputi:
  - Perancangan, terdiri dari perancangan *database*, antarmuka sistem, serta alat-alat apa saja yang diperlukan dalam pengambilan data.
  - Pengkodean, mengimplementasikan teori-teori serta metode-metode yang digunakan kedalam bentuk program dengan bahasa pemrograman PHP.
  - Pengujian, melakukan pengujian sebagai suatu proses pembuktian dari sistem yang sudah dibuat.

## 1.6 Sistematika Penulisan

Sistematika yang digunakan penulis dalam menulis laporan penelitian ini adalah sebagai berikut :

BAB 1 Pendahuluan merupakan pengantar untuk menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian dan sistematika penulisan yang digunakan oleh penulis.

BAB 2 Tinjauan Pustaka menjelaskan konsep yang mendasari penelitian dan landasan teori yang akan digunakan.

BAB 3 Analisis dan Perancangan Sistem membahas mengenai gambaran umum proses yang terjadi dalam pengolahan data, rancangan *database*, pembuatan diagram aliran data pada sistem, perancangan tampilan masukan dan keluaran sistem.

BAB 4 Implementasi dan Analisis Sistem berisi tentang penerapan program dan cara kerja sistem yang telah dibangun.

BAB 5 Kesimpulan dan Saran merupakan hasil dari implementasi sistem mengenai kinerja sistem secara keseluruhan, yang dijabarkan dalam bentuk kesimpulan dan saran oleh penulis.

©UKDW



## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan Hasil analisis pembentukan pola asosiasi kumpulan kata kunci, maka didapatkan kesimpulan sebagai berikut:

1. Sistem yang dibuat mampu membentuk kumpulan kata kunci menjadi pola-pola yang asosiatif. Namun, ketika diujikan ke data uji nilai *confidence* dari pola asosiasi yang terbentuk banyak yang tidak akurat.
2. Pembentukan pola asosiasi kumpulan kata kunci dengan teknik *association rule* sangat dipengaruhi oleh sumber data dan pada hasil pembentukan kata kunci itu sendiri.
3. Hasil pembentukan pola asosiasi kata kunci dengan teknik *association rule* sangat tergantung dari besarnya nilai *support* dan *confidence*.

#### 5.2 Saran

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

1. Hasil pembentukan pola asosiasi kata kunci akan jauh lebih optimal jika sumber data yang dianalisis adalah dokumen karya ilmiah pada bab 2 atau bab 3 karena cenderung banyak menggunakan istilah-istilah teknik.
2. Hasil pembentukan pola asosiasi kata kunci akan jauh lebih bermanfaat apabila kata kunci yang digunakan adalah kata kunci yang terdiri lebih dari dua kata.

## DAFTAR PUSTAKA

- Erwin. (2009). *Analisis Market Basket dengan Algoritma Apriori dan FP-Growth*. diakses dari: [www.uppm.ilkom.unsri.ac.id/userfiles/JurnalVol\\_4\\_No\\_2\\_Juli\\_2009/6-Erwin.pdf](http://www.uppm.ilkom.unsri.ac.id/userfiles/JurnalVol_4_No_2_Juli_2009/6-Erwin.pdf)
- Feldman, R., Sanger, J. (2007). *The Text Mining Handbook Advanced Approaches in Analysing Unstructured Data*.
- Gunarto, T. Y. (2009). *Uji Chi Kuadrat*. Dapat diakses di: [www.thomasyg.staff.gunadarma.ac.id/download/files/8193/Uji+Chi+Square.pdf](http://www.thomasyg.staff.gunadarma.ac.id/download/files/8193/Uji+Chi+Square.pdf)
- Han, J., Kamber, M. (2001). *Data Mining Concepts and Techniques*.
- Kantardzic M. (2003). *Data Mining Concepts, Models, Methods and Algorithms*.
- Mahgoub, H. (2006). *Mining Association Rules from Unstructured Data*. World Academy of Science, Engineering and Technology 20. dapat diakses di: [www.waset.org/journals/waset/v20/v20-1.pdf](http://www.waset.org/journals/waset/v20/v20-1.pdf)
- Pal, S. (2009). *Finding Phrases Two Statistical Approaches*. dapat diakses di: [www.sujitpal.blogspot.com/2009/11/finding-phrases-two-statistical.html](http://www.sujitpal.blogspot.com/2009/11/finding-phrases-two-statistical.html).