

**PENERAPAN ALGORITMA TF-IDF DAN ALGORITMA VECTOR  
SPACE MODEL DALAM PENCARIAN JURNAL BERBASIS ONLINE  
STUDI KASUS : PERPUSTAKAAN UKDW**

Skripsi



oleh  
**ADITYA SATRIO KUMOLO**  
**23090482**

PROGRAM STUDI SISTEM INFORMASI FAKULTAS TEKNOLOGI  
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA  
TAHUN 2013

**PENERAPAN ALGORITMA TF-IDF DAN ALGORITMA VECTOR  
SPACE MODEL DALAM PENCARIAN JURNAL BERBASIS ONLINE  
STUDI KASUS : PERPUSTAKAAN UKDW**

Skripsi



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

Disusun oleh

**ADITYA SATRIO KUMOLO**  
**23090482**

PROGRAM STUDI SISTEM INFORMASI FAKULTAS TEKNOLOGI  
INFORMASI UNIVERSITAS KRISTEN DUTA WACANA  
TAHUN 2013

## HALAMAN PERSETUJUAN

Judul Skripsi : Penerapan Algoritma TF-IDF dan Algoritma Vector  
Space Model dalam Pencarian Jurnal Berbasis Online  
Studi Kasus : Perpustakaan UKDW

Nama Mahasiswa : ADITYA SATRIO KUMOLO

N I M : 23090482

Matakuliah : Skripsi

Kode : SI4046

Semester : Genap

Tahun Akademik : 2012/2013

Telah diperiksa dan disetujui di Yogyakarta,  
Pada tanggal 15 Agustus 2013

Dosen Pembimbing I



YETLI OSLAN, S.Kom., M.T.

Dosen Pembimbing II



ANTONIUS RACHMAT C, S.Kom., M.Cs

## HALAMAN PENGESAHAN

### PENERAPAN ALGORITMA TF-IDF DAN ALGORITMA VECTOR SPACE MODEL DALAM PENCARIAN JURNAL BERBASIS ONLINE STUDI KASUS : PERPUSTAKAAN UKDW

Oleh: ADITYA SATRIO KUMOLO / 23090482

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Sistem Informasi Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal  
31 Juli 2013

Yogyakarta, 15 Agustus 2013  
Mengesahkan,

Dewan Penguji:

1. YETLI OSLAN, S.Kom., M.T.
2. ANTONIUS RACHMAT C, S.Kom., M.Cs
3. UMI PROBOYEKTI, S.Kom., MLIS.
4. Ir. NJOO HARIANTO KRISTANTO, M.T., M.M.

**DUTA WACANA**

Dekan  
  
(Drs. WIMMIE HANDIWIDJOJO, M.T.)

Ketua Program Studi  
  
(YETLI OSLAN, S.Kom., M.T.)

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

**Penerapan Algoritma TF-IDF dan Algoritma Vector Space Model dalam  
Pencarian Jurnal Berbasis Online  
Studi Kasus : Perpustakaan UKDW**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Sistem Informasi 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 hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 15 Agustus 2013



**ADITYA SATRIO KUMOLO**

23090482



## FORMULIR PERBAIKAN (REVISI) SKRIPSI

Dicetak tanggal: 15-08-2013 15:49:16

Yang bertanda tangan di bawah ini:

Nama : ADITYA SATRIO KUMOLO

N I M : 23090482

Judul Skripsi : PENERAPAN ALGORITMA TF-IDF DAN ALGORITMA VECTOR SPACE MODEL  
DALAM PENCARIAN JURNAL BERBASIS ONLINE  
STUDI KASUS : PERPUSTAKAAN UKDW

Tanggal Pendadaran : Rabu, 31 Juli 2013 pukul 10:00 WIB

Telah melakukan perbaikan tugas akhir dengan lengkap.

Demikian pernyataan kami agar dapat dipergunakan sebagaimana mestinya.

Yogyakarta, Kamis, 15 Agustus 2013

Dosen Pembimbing I

YETLI OSLAN, S.Kom., M.T.

Dosen Pembimbing II

ANTONIUS RACHMAT C, SKom.,M.Cs

## UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yesus Kristus YME atas berkat yang diberikan-Nya sehingga penulis dapat menyelesaikan tugas akhir yang ditempuh dengan lancar. Selain itu penulis juga mengucapkan terima kasih kepada :

1. Ibu Yetli Oslan sebagai pembimbing I yang telah sabar dalam membimbing penulis dari awal hingga akhir proses pembuatan tugas akhir.
2. Bapak Antonius Rachmat sebagai pembimbing II yang telah memberikan banyak masukan dan ide untuk menyelesaikan tugas akhir ini.
3. Bapak Njoo Harianto Kristanto sebagai penguji yang telah memberikan koreksi untuk kesalahan yang terlewatkan oleh penulis.
4. Ibu Umi Proboyekti sebagai penguji dan sebagai wali dosen selama lebih dari tiga tahun yang telah membimbing penulis.
5. Kedua orang tua dan kakak penulis yang selalu memberikan dorongan agar dapat menyelesaikan kuliah dengan baik.
6. Seluruh Bapak/Ibu dosen dan staf Program Studi Sistem Informasi yang telah memberikan ilmu pengetahuan dan informasi yang berguna selama kegiatan perkuliahan hingga dapat diterapkan pada dunia kerja.
7. Teman-teman Sistem Informasi angkatan 2009 yang selalu mendukung penulis untuk menyelesaikan tugas akhir ini.
8. Teman-teman mahasiswa lain baik dari UKDW maupun universitas lain yang telah memberikan dukungan untuk menyelesaikan tugas akhir ini.
9. Liem Lina yang selalu mendoakan, mendukung, menyemangati, dan memaksa penulis agar dapat menyelesaikan tugas akhir dengan baik.

Demikian ucapan terima kasih yang dapat penulis berikan. Semoga Tuhan YME selalu memberkati apa yang penulis maupun teman-teman kerjakan.

## ABSTRAK

Perpustakaan merupakan tempat yang penting di dalam sebuah universitas karena menyimpan berbagai informasi dan pengetahuan yang sangat banyak dan bermanfaat. Bagi mahasiswa maupun pengguna perpustakaan lainnya dapat mencari koleksi periodikal yang sesuai dengan kebutuhannya di perpustakaan. Namun dikarenakan banyaknya periodikal yang dimiliki oleh perpustakaan, pengguna perpustakaan membutuhkan waktu yang lama dalam melakukan pencarian dengan membaca satu persatu isi dari periodikal tersebut.

Berdasarkan permasalahan diatas, dibangunlah sistem pencarian periodikal menggunakan metode *Term Frequency Inverse Document Frequency* dan *Vector Space Model* yang digunakan untuk memberikan bobot terhadap periodikal yang akan dicari oleh pengguna. Selain itu *feature selection* diterapkan untuk mendukung metode TF-IDF dan VSM agar dapat berjalan dengan lebih cepat dan efisien. Dengan pembobotan yang dilakukan terhadap periodikal, diharapkan periodikal yang akan disajikan akan sesuai dengan apa yang dicari pengguna.

Dengan pemanfaatan *feature selection* dalam mendukung pembobotan periodikal tetap dapat memberikan tingkat akurasi yang tinggi mencapai 100% dan mempercepat waktu pencarian dalam sistem pencari. Dan dengan memodifikasi langkah pencarian yang umum digunakan dapat memberikan hasil yang lebih baik dari segi waktu yang mencapai 30 (tiga puluh) kali lebih cepat.

**Kata kunci** : Term Frequency – Inverse Document Frequency (tf-idf), Vector Space Model, Feature Selection, Precision and Recall, Periodikal, Information Retrieval.



## DAFTAR ISI

HALAMAN SAMPUL DALAM .....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN KARYA .....	v
HALAMAN PERBAIKAN (REVISI) SKRIPSI.....	vi
UCAPAN TERIMAKASIH.....	vii
ABSTRAK.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
DAFTAR GRAFIK.....	xiv
DAFTAR SINTAK DAN PSEUDO.....	xv
BAB I.....	1
PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	1
1.3 Batasan Masalah.....	1
1.4 Spesifikasi Sistem.....	2
1.5 Tujuan dan Manfaat Penelitian.....	3
1.6 Metodologi Penulisan .....	3
1.7 Sistematika Penulisan.....	4
BAB II.....	6
LANDASAN TEORI.....	6
2.1 Tinjauan Putaka .....	6
2.2 Landasan Teori .....	7
2.2.1 Perpustakaan .....	7

2.2.2	Koleksi .....	7
2.2.3	Text Mining.....	8
2.2.4	Information Retrieval .....	8
2.2.5	Tokenizing.....	9
2.2.6	Stopword .....	10
2.2.7	Stemming .....	10
2.2.8	Feature Selection .....	11
2.2.9	Term Frequency – Inverse Document Frequency .....	12
2.2.10	Vector Space Model .....	13
2.2.11	Precision And Recall.....	14
BAB III .....		15
ANALISIS DAN RANCANGAN .....		15
3.1	Rancangan Sistem .....	15
3.1.1	Use Case Diagram.....	15
3.1.2	Data Flow Diagram (DFD).....	15
3.1.3	Model Data Logika (MDL).....	17
3.1.4	Pra Proses .....	23
3.1.5	Proses TF-IDF.....	25
3.1.6	Proses Pencarian.....	26
3.1.7	Proses VSM.....	28
3.2	Rancangan Hasil Sistem.....	29
3.2.1	Rancangan Search Engine.....	29
3.2.2	Rancangan Output.....	29
3.3	Pengujian Sistem .....	30
3.3.1	Menghitung Precision Dan Recall .....	30
3.3.2	Menguji Pengaruh Presentase Feature Selection .....	31
3.3.3	Menguji Pengaruh Perbedaan Langkah Pencarian.....	31
BAB IV .....		32
PENERAPAN DAN ANALISIS SISTEM .....		32
4.1.	Implementasi Sistem .....	32
4.1.1.	Pra Proses .....	32

4.1.2.	Form Utama .....	33
4.1.3.	Form Output .....	33
4.1.4.	Form Penelitian <i>Precision And Recall</i> .....	35
4.2.	Analisis Sistem .....	35
4.2.1.	Proses Pencarian.....	35
4.3.	Pengujian Sistem .....	38
4.3.1.	Pengujian Pertama.....	38
4.3.1.1.	<i>Feature Selection</i> 10% .....	39
4.3.1.2.	<i>Feature Selection</i> 20% .....	41
4.3.1.3.	<i>Feature Selection</i> 30% .....	44
4.3.1.4.	<i>Feature Selection</i> 40% .....	47
4.3.1.5.	<i>Feature Selection</i> 50% .....	50
4.3.1.6.	Hasil Keseluruhan Pengujian <i>Feature Selection</i> .....	53
4.3.2.	Pengujian Kedua .....	54
BAB V	.....	56
KESIMPULAN DAN SARAN	.....	56
5.1.	Kesimpulan.....	56
5.2.	Saran .....	56
DAFTAR PUSTAKA	.....	57
LAMPIRAN A	.....	A-1
LAMPIRAN B	.....	B-1

## DAFTAR GAMBAR

Gambar 2.1 Retrieving Matched Document .....	8
Gambar 2.2 Tokenizing Algorithm .....	9
Gambar 2.3 Stemming Algorithm.....	11
Gambar 3.1 Use Case Diagram.....	15
Gambar 3.2 Diagram Konteks.....	16
Gambar 3.3 DFD Level 0.....	16
Gambar 3.4 DFD Level 1 Pencarian .....	17
Gambar 3.5 MDL 1 .....	17
Gambar 3.6 MDL 2 .....	18
Gambar 3.7 MDL 3 .....	18
Gambar 3.8 MDL 4 .....	19
Gambar 3.9 MDL 6.....	20
Gambar 3.10 <i>Flowchart</i> Pra Proses Pertama .....	23
Gambar 3.11 <i>Flowchart</i> Pra Proses Kedua.....	24
Gambar 3.12 <i>Flowchart</i> TF-IDF .....	25
Gambar 3.13 <i>Flowchart</i> Pencarian Yang Standart.....	26
Gambar 3.14 <i>Flowchart</i> Pencarian Setelah Dimodifikasi .....	27
Gambar 3.15 <i>Flowchart</i> VSM .....	28
Gambar 3.16 Tampilan Halaman <i>Search</i> .....	29
Gambar 3.17 Tampilan Hasil Pencarian .....	29
Gambar 3.18 Tampilan Form Pengujian <i>Precision And Recall</i> .....	30
Gambar 4.1 Form Utama.....	33
Gambar 4.2 Form Output .....	34
Gambar 4.3 Form Penelitian <i>Precision and Recall</i> .....	35
Gambar 4.4 Hasil Pencarian Menggunakan Langkah Pertama.....	54
Gambar 4. 5 Hasil Pencarian Menggunakan Langkah Kedua .....	55

## DAFTAR TABEL

Tabel 2.1 Tabel Perhitungan <i>TF-IDF</i> (Herwansyah, 2012).....	13
Tabel 2.2 Tabel Perhitungan <i>Vector Space Model</i> (Herwansyah, 2012) .....	14
Tabel 3.1 Kunci Aturan Bisnis.....	19
Tabel 3.2 Entitas tblartikel .....	21
Tabel 3.3 Entitas tblkata.....	21
Tabel 3.4 Entitas tblcosine .....	22
Tabel 3.5 Entitas stopword.....	22
Tabel 3.6 Entitas stemming.....	22
Tabel 4.1 Hasil Pencarian Dengan FS 10% .....	39
Tabel 4.2 Tabel Hasil Perhitungan <i>Precision and Recall</i> FS 10% .....	41
Tabel 4.3 Hasil Pencarian Dengan FS 20% .....	41
Tabel 4.4 Tabel Hasil Perhitungan <i>Precision and Recall</i> FS 20% .....	44
Tabel 4.5 Hasil Pencarian Dengan FS 30% .....	44
Tabel 4.6 Tabel Hasil Perhitungan <i>Precision and Recall</i> FS 30% .....	47
Tabel 4.7 Hasil Pencarian Dengan FS 40% .....	47
Tabel 4. 8 Tabel Hasil Perhitungan <i>Precision and Recall</i> FS 40% .....	50
Tabel 4.9 Hasil Pencarian Dengan FS 50% .....	50
Tabel 4.10 Tabel Hasil Perhitungan <i>Precision and Recall</i> FS 50% .....	53

## DAFTAR GRAFIK

Grafik 4.1 Grafik Cosine Hasil Pencarian Dengan FS 10% .....	40
Grafik 4.2 Grafik <i>Precision and Recall</i> Dengan FS 10% .....	41
Grafik 4.3 Grafik Cosine Hasil Pencarian Dengan FS 20% .....	43
Grafik 4.4 Grafik <i>Precision and Recall</i> Dengan FS 20% .....	44
Grafik 4.5 Grafik Cosine Hasil Pencarian Dengan FS 30% .....	46
Grafik 4.6 Grafik <i>Precision and Recall</i> Dengan FS 30% .....	47
Grafik 4.7 Grafik Cosine Hasil Pencarian Dengan FS 40% .....	49
Grafik 4.8 Grafik <i>Precision and Recall</i> Dengan FS 40% .....	50
Grafik 4.9 Grafik Cosine Hasil Pencarian Dengan FS 50% .....	52
Grafik 4.10 Grafik <i>Precision and Recall</i> Dengan FS 50% .....	53

©UKYDOWN

## DAFTAR SINTAK DAN PSEUDO CODE

Kode 4.1 Sintak <i>Tokenizing</i> Kata Kunci .....	36
Pseudo Code 4.1 Pseudo <i>Tokenizing</i> Kata Kunci .....	36
Kode 4.2 Sintak Pencarian Artikel.....	37
Pseudo Code 4.2 Pseudo Pencarian Artikel .....	38

©UKDW

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Perpustakaan UKDW merupakan salah satu perpustakaan yang memiliki koleksi kumpulan artikel dari berbagai bidang akademik. Sumber dari artikel tersebut tidak hanya didapat dari pihak internal UKDW, melainkan didapat dari universitas lain dengan cara berlangganan. Akan tetapi, perpustakaan UKDW sekarang ini masih belum memiliki sistem yang digunakan untuk penyimpanan data-data artikel dan pencarian artikel. Dengan tidak adanya sistem penyimpanan dan pencarian artikel, pengunjung tidak dapat mencari artikel secara efektif dan tidak mengetahui apakah perpustakaan UKDW memiliki artikel yang dikehendaki pengunjung. Oleh karena itu, penelitian ini bermaksud untuk memecahkan permasalahan tersebut dengan membuat sistem untuk penyimpanan artikel serta sistem temu kembali informasi untuk artikel menggunakan algoritma TF-IDF dan Vector Space Model.

### 1.2 Rumusan Masalah

Bagaimana menerapkan algoritma *TF-IDF* dan *Vector Space Model* dalam memberikan bobot nilai untuk setiap dokumen yang dicari sesuai dengan kata kunci yang dimasukkan?

Dari *feature selection* sebesar 10%, 20%, 30%, 40%, dan 50% manakah yang memberikan ketepatan hasil pencarian lebih baik dan berapakah nilai *precision recall* yang dihasilkan masing-masing *feature selection*?

### 1.3 Batasan Masalah

Batasan-batasan yang ditentukan dalam penelitian ini antara lain :

- a. Artikel yang digunakan merupakan artikel koleksi Perpustakaan UKDW.
- b. Algoritma yang digunakan dalam pencarian adalah algoritma *Term Frequency Inverse Document Frequency* (TF-IDF) dan *Vector Space Model* (VSM).



- c. Pencarian terbatas pada judul, abstrak dan kata kunci.
- d. Proses stemming hanya dilakukan untuk *token* berbahasa inggris.
- e. *Stopword List* yang digunakan didapat dari <http://www.textfixer.com/resources/common-english-words.txt> dan [http://web.archive.org/web/20100104090147/http://fpmipa.upi.edu/staff/yudi/stop\\_words\\_list.txt](http://web.archive.org/web/20100104090147/http://fpmipa.upi.edu/staff/yudi/stop_words_list.txt)
- f. Kata *stemming* bahasa inggris diambil dari <http://tartarus.org/~martin/PorterStemmer/voc.txt> dan <http://tartarus.org/~martin/PorterStemmer/output.txt>

## 1.4 Spesifikasi Sistem

- a. Spesifikasi Program
  - 1) admin
    - a) Program dapat digunakan untuk mengatur setup yang berhubungan dengan artikel, *stopword*, dan *stemming*.
  - 2) user
    - a) Program mampu mengolah kata kunci yang dimasukkan pengguna untuk dapat diolah dalam pencarian artikel yang sesuai dengan kebutuhan pengguna.
    - b) Program mampu menampilkan hasil pencarian yang sudah diurutkan berdasarkan tingkat similaritasnya terhadap kata kunci.
- b. Spesifikasi Hardware
  - 1) *Harddisk* dengan kapasitas minimal 64 Gb
  - 2) *RAM* minimal 512 Mb
  - 3) *Processor* minimal Pentium 4
  - 4) Monitor minimal 14 inch
  - 5) *Keyboard* dan *Mouse*
  - 6) Koneksi internet
- c. Spesifikasi Software
  - 1) *Browser* (Mozilla Firefox, Google Chrome / Chrome Canary / Chromium, Safari, Opera, Maxthon)
  - 2) Sistem Operasi Microsoft® Windows® Xp atau 7

d. Spesifikasi Brainware

- 1) Mampu menggunakan komputer, browser serta mengakses ke internet

## 1.5 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah untuk membuktikan algoritma TF-IDF dan VSM mampu mengukur tingkat similaritas artikel yang akan dicari berdasarkan kata kunci yang dimasukkan pengguna. Semakin besar tingkat similaritas tersebut, maka semakin cocok artikel tersebut dengan kata kunci.

## 1.6 Metodologi Penulisan

a. Studi Literatur

Tahap ini merupakan tahap awal penelitian dimana penulis mempelajari cara kerja metode *TF-IDF* dan *VSM* dengan mengumpulkan berbagai informasi yang mampu memperjelas proses serta cara kerja metode tersebut. Selain mempelajari metode *TF-IDF* dan *VSM*, penulis juga mempelajari beberapa proses yang mendukung ke dua metode tersebut seperti *tokenizing*, *filtering* dan *stemming*.

b. Pengumpulan Data

Terdapat beberapa data yang dibutuhkan dalam penelitian ini, data pertama adalah data jurnal periode 2009–2012 yang dimiliki perpustakaan UKDW. Jurnal yang akan digunakan adalah jurnal yang terbatas pada dua bahasa, yaitu bahasa Indonesia dan bahasa Inggris. Setelah mengumpulkan data jurnal, maka data selanjutnya yang harus dikumpulkan adalah data *stopword* yang dapat diunduh pada <http://www.textfixer.com/resources/common-english-words.txt> dan [http://web.archive.org/web/20100104090147/http://fpmipa.upi.edu/staff/yud/i/stop\\_words\\_list.txt](http://web.archive.org/web/20100104090147/http://fpmipa.upi.edu/staff/yud/i/stop_words_list.txt). Kemudian data terakhir yang wajib dimiliki adalah data yang digunakan untuk proses *stemming*, data tersebut didapat dengan mengunduh pada <http://tartarus.org/~martin/PorterStemmer/voc.txt> dan <http://tartarus.org/~martin/PorterStemmer/output.txt>.

c. Perancangan Sistem

Setelah semua data diperoleh, maka tahap selanjutnya adalah merancang database, antarmuka serta proses pencarian.

#### d. Pembangunan Sistem

Tahap akhir yang akan dilakukan adalah pembuatan *search engine* yang akan melakukan pencarian jurnal berdasarkan judul, abstrak, dan *keyword*. Kata kunci yang dimasukkan akan mengalami proses *tokenizing* yang bertujuan memotong kata kunci tersebut menjadi pecahan-pecahan kata yang membentuk kalimat kata kunci tersebut. Kemudian hasil dari *tokenizing* tersebut akan memasuki proses *filtering* yang bertujuan mengambil kata-kata penting dan membuang kata yang kurang penting yang terdapat pada *stopword*.

Setelah didapat kumpulan kata hasil *filtering*, proses *stemming* akan dilakukan. Tujuan dari *stemming* adalah mengubah kata tersebut menjadi kata dasar. Dan proses terakhir adalah *analyzing*, yaitu menentukan keterhubungan antara kata kunci yang dimasukkan dengan atribut-atribut yang melekat pada jurnal.

### 1.7 Sistematika Penulisan

Bab I yang merupakan bagian Pendahuluan terdiri dari beberapa bagian. Bagian pertama yang merupakan latar belakang masalah berisi mengenai permasalahan yang timbul serta solusi pemecahan masalah tersebut. Rumusan masalah merupakan bagian dimana permasalahan yang menjadi inti penelitian dituliskan. Batasan masalah berisi tentang jangkauan dari penelitian yang dilakukan supaya fokus dari permasalahan penelitian dapat lebih jelas. Spesifikasi sistem berisi tentang piranti lunak atau keras apa saja yang dibutuhkan dalam penelitian ini. Tujuan penelitian tidak hanya berisi tentang penyelesaian masalah, namun pada bagian ini berisi tentang pengembangan ilmu bagi para pengguna. Metodologi penelitian berisi langkah-langkah yang dilakukan oleh penulis dalam menyelesaikan suatu permasalahan. Sistematika penulisan merupakan bagian dimana rancangan laporan yang akan dibuat dituliskan. Tujuan dari bagian ini adalah memberikan penjelasan singkat dari tiap bab supaya lebih mudah dibaca oleh pengguna.

Bab II yang merupakan bagian Landasan Teori ini terdiri dari teori-teori yang mendasari dan mendukung penelitian. Selain teori terdapat metode TF-IDF dan VSM yang digunakan dalam penelitian ini. Contoh-contoh sederhana dalam penggunaan metode dapat ditemukan pada bab ini, hal tersebut bertujuan untuk memberikan gambaran sederhana agar pengguna mudah memahami cara kerja dari metode yang digunakan.

Bab III yang merupakan bagian Analisis dan Rancangan ini terdiri dari segala bentuk pola pikir penulis yang digambarkan dalam berbagai macam bentuk rancangan. Terdapat beberapa rancangan seperti rancangan database, rancangan tampilan, serta rancangan sistem yang dijelaskan dalam bentuk *flowchart*, DFD (*Data Flow Diagram*) dan *use case*. Bab ini akan menggunakan banyak tampilan visual agar lebih mudah dipahami bagi pengguna.

Bab IV yang merupakan bagian Penerapan dan Analisis Sistem akan menjabarkan hasil dari sistem yang telah dibuat. Terdapat tampilan-tampilan *form* yang dihasilkan dari program tersebut. Selain itu, bagian ini akan menampilkan hasil dari proses perhitungan metode TF-IDF dan VSM.

Bab V adalah bagian Penutup. Pada bagian ini terdiri dari kesimpulan dan saran dari penulis. Kesimpulan tersebut merupakan jawaban dari rumusan masalah yang telah ditulis pada Bab I. Saran yang terdapat pada bagian ini merupakan saran pengembangan yang dapat dimanfaatkan untuk melanjutkan penelitian yang serupa.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan hasil analisis dan pengujian sistem, maka dapat disimpulkan sebagai berikut :

- a. *Feature selection* sebesar 10% mampu memberikan tingkat *precision* sebesar 100% sedangkan *feature selection* yang lain hanya dapat memberikan tingkat *precision* berkisar 80%.
- b. Untuk melakukan pencarian artikel yang sangat banyak disarankan menggunakan *feature selection* 30% atau 40%.
- c. Dengan menggunakan langkah pencarian yang sudah dimodifikasi, waktu yang digunakan sistem untuk melakukan pembobotan lebih cepat 30 (tiga puluh) kali dibandingkan langkah pencarian yang umum digunakan.

#### 5.2. Saran

Saran yang diberikan untuk mengembangkan sistem pencarian ini adalah :

- a. Menambahkan data untuk *stemming* dalam bahasa Indonesia, dikarenakan data *stemming* yang ada sekarang ini dalam bahasa Inggris.
- b. Mengembangkan langkah pencarian agar dapat berjalan lebih cepat dari yang sudah dibuat.

## DAFTAR PUSTAKA

- , *Teaching & Learning: What is a Periodical?* (2010). Retrieved April 8, 2013, from Indiana University Bloomington: <http://www.libraries.iub.edu/?pageId=1002225>
- Aggarwal, C. C., & Zhai, C. (2012). *Mining Text Data*. New York: Springer Science+Business Media.
- Ariafandi, F. (2011). *Implementasi Vector Space Model Untuk Pencarian Ayat Dalam Kitab Perjanjian Baru*. (Undergraduate thesis, Duta Wacana Christian University, 2011). Retrieved from <http://sinta.ukdw.ac.id>
- Herwansyah, A. (2012). APLIKASI PENGKATEGORIAN DOKUMEN DAN PENGUKURAN. Skripsi Sistem Informasi, Fakultas Ilmu Komputer Universitas Gunadarma.
- Lie, N.M. (1997). *Implementasi Pemrograman Berorientasi Obyek Untuk Program Penyediaan Informasi Perpustakaan*. (Undergraduate thesis, Duta Wacana Christian University, 1997). Retrieved from <http://sinta.ukdw.ac.id>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Feature selection*. Retrieved April 7, 2013, from stanford.edu: <http://nlp.stanford.edu/IR-book/html/htmledition/feature-selection-1.html>
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge.
- Nugroho, R.W.T. (2006). *Program Bantu Peringkasan Dokumen Teks Berbahasa Inggris Dengan Algoritma Porter Stemming*. (Undergraduate thesis, Duta Wacana Christian University, 2006). Retrieved from <http://sinta.ukdw.ac.id>
- Tiatira, R. (2009). *Implementasi Vector Space Model Pada Aplikasi Mp3 Search Online*. (Undergraduate thesis, Duta Wacana Christian University, 2009). Retrieved from <http://sinta.ukdw.ac.id>
- Wess, S. M., Indurkha, N., Zhang, T., & Damerau, F. J. (2005). *Text Mining Predictive Methods for Analyzing Unstructured Information*. United States of America: Springer.