

# **Implementasi Vector Space Model Untuk Pencarian Ayat dalam Kitab Perjanjian Baru**

TUGAS AKHIR



Disusun oleh :

**FRANDANIS ARIAFANDI**

**NIM.22033312**

**Program Studi Teknik Informatika**



**Program Studi Teknik Informatika Fakultas Teknik**

**Universitas Kristen Duta Wacana**

**2011**

# **Implementasi Vector Space Model Untuk Pencarian Ayat dalam Kitab Perjanjian Baru**

**TUGAS AKHIR**



**Diajukan kepada Fakultas Teknik Program Studi Teknik Informatika  
Universitas Kristen Duta Wacana**

**Sebagai salah satu syarat dalam memperoleh gelar  
Sarjana Komputer**

**Disusun oleh :**

**FRANDANIS ARIAFANDI**

**NIM.22033312**

**Program Studi Teknik Informatika Fakultas Teknik  
Universitas Kristen Duta Wacana**

**2011**

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul:

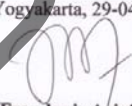
IMPLEMENTASI VECTOR SPACE MODEL UNTUK PENCARIAN

AYAT DALAM KITAB PERJANJIAN BARU

yang saya kerjakan untuk melengkapi 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 adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia menerima sanksi berupa pencabutan gelar kesarjanaan saya.

Yogyakarta, 29-04-2011

  
(Frandanis Ariafandi)

22033312



**HALAMAN PERSETUJUAN**


Judul : IMPLEMENTASI VECTOR SPACE MODEL UNTUK  
PENCARIAN AYAT DALAM KITAB PERJANJIAN BARU  
Nama : Frandanis Ariafandi  
NIM : 22033312  
Mata Kuliah : Tugas Akhir Kode : TI2126  
Semester : Genap Tahun Akademik : 2010/2011

Telah diperiksa dan disetujui  
di Yogyakarta,  
pada tanggal 29-04-2011

Dosen Pembimbing I

  
Lucia Dwi Krisnawati, M.A.

Dosen Pembimbing II

  
Restyandito, S.Kom, MSIS.

HALAMAN PENGESAHAN

SKRIPSI

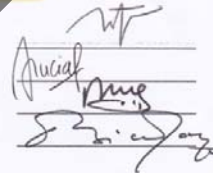
IMPLEMENTASI VECTOR SPACE MODEL UNTUK PENCARIAN AYAT  
DALAM KITAB PERJANJIAN BARU  
Oleh: Frandanis Ariafandi/22033312

Dipertahankan di depan dewan Penguji Tugas Akhir/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

Yogyakarta, 20 Mei 2011  
Mengesahkan,

Dewan Penguji:

1. Restyandito, S.Kom, MSIS.
2. Lucia Dwi Krisnawati, M.A.
3. Nugroho Agus. H. S.Si, M.Si.
4. Drs. Wimmie Handiwidjojo, MIT.

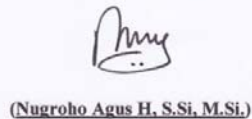


Dekan

Ketua Program Studi



(Drs. Wimmie Handiwidjojo, MIT.)



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

## UCAPAN TERIMAKASIH

Segala puji syukur penulis ucapkan kepada Tuhan Yesus yang selalu memberi harapan dalam hidup penulis dan yang telah memberikan semua pihak yang terlibat dalam pembuatan program dan penyusunan laporan ini.

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

Dalam menyelesaikan pembuatan program dan laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, saran dan masukan dari berbagai pihak, baik secara langsung maupun secara tidak langsung. Untuk itu dengan segala kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. [Lucia DwiKrisnawati, M.A.](#) selaku dosen pembimbing I yang telah memberikan bimbingannya dengan sabar dan baik kepada penulis, juga kepada
2. **Restyandito, S. Kom, MSIS.** selaku dosen pembimbing II atas bimbingan, petunjuk dan masukan yang diberikan selama pengerjaan tugas ini sejak awal hingga akhir.
3. Keluarga tercinta yang memberikan dukungan dan semangat, Bapak-Ibu, dan Yohana.
4. Orang-orang terdekat yang telah memberikan dukungan dan semangat, LIFE.Com, undrGod777, '03 TI ukdw
5. Pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa program dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca sekalian, sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan maupun yang pernah penulis lakukan sewaktu membuat program Tugas Akhir. Dan semoga ini dapat berguna bagi kita semua.

Yogyakarta, April 2011

Penulis

© UKDW

## Abstrak

Salah satu sumber utama dalam mempelajari agama Kristen adalah Alkitab. Alkitab terdiri dari dua bagian yaitu Kitab Perjanjian Lama dan Kitab Perjanjian Baru. Kebanyakan orang Kristen merenungkan Alkitab setiap hari. Namun orang Kristen sering kali mengalami kesulitan untuk menemukan ayat-ayat yang mereka butuhkan pada saat tertentu. Pencarian secara manual membutuhkan waktu yang lama. Permasalahan tersebut dapat diselesaikan dengan Sistem Temu Kembali Informasi. Karena Alkitab merupakan dokumen yang sangat besar maka penulis membatasi sistem dengan pencarian hanya dalam Kitab Perjanjian Baru. Sistem Temu Kembali Informasi terdiri atas dua bagian, yaitu penyimpanan dokumen dan pemrosesan *query*. Salah satu metode untuk merepresentasikan Sistem Temu Kembali Informasi adalah *Vector Space Model (VSM)*. *Vector Space Model* merupakan kombinasi metode *TF-IDF* dengan *Cosine*. Metode *TF-IDF* digunakan untuk menghitung bobot token. Proses yang dilakukan terhadap dokumen secara berurutan adalah tokenisasi, stopword, penghitungan *TF-IDF*, kemudian disimpan dalam file. *Query* yang dimasukkan pengguna dicari nilai *TF-IDF* kemudian dihitung nilai *Cosine* terhadap dokumen. *Cosine* adalah ukuran kemiripan dokumen.

Hasil pengujian dari implementasi *Vector Space Model* untuk sistem temu kembali ayat Kitab Perjanjian Baru menggunakan *query* satu kata menghasilkan nilai presisi satu, representatif dari nilai tertinggi. Pengujian menggunakan dua kata menghasilkan nilai presisi yang cenderung menurun.

Kata kunci : *Vector Space Model (VSM)*, *TF-IDF*, Presisi, *Recall*.



## DAFTAR ISI

Halaman Judul.....	
Halaman Keaslian Skripsi.....	i
Halaman Persetujuan .....	ii
Halaman Pengesahan .....	iii
Ucapan Terima Kasih.....	iv
Abstraksi .....	vi
Daftar Isi .....	vii
Daftar Tabel .....	ix
Daftar Gambar .....	x
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	2
1.5 Metode Penelitian .....	2
1.6 Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Tinjauan Pustaka.....	4
2.2 Landasan Teori.....	7
2.2.1 Sistem Temu Kembali Informasi .....	7
2.2.2 Metode Ruang Vector ( <i>Vector Space Model</i> ).....	9
2.2.2.1 Pengindeksan Dokumen.....	12
2.2.2.2 <i>Term Weighting</i> (Pembobotan Term).....	13
2.2.2.3 Ukuran Kesamaan .....	16
BAB III PERANCANGAN SISTEM .....	20
3.1 Alat dan Bahan.....	20

3.2 Perancangan Antar Muka.....	20
3.3 Flowchart .....	22
3.3.1 Flowchart Program Utama .....	22
3.3.2 Flowchart Pra Proses.....	23
3.3.3 Flowchart Pembobotan Dokumen.....	24
3.3.4 Flowchart Pemrosesan Kata Kunci .....	25
3.3.5 Flowchart Pencarian.....	26
BAB IV IMPLEMENTASI DAN ANALISIS SISTEM .....	27
4.1 Implementasi Sistem .....	27
4.2 Implementasi <i>Vector Space Model</i> .....	30
4.2.1 Tokenisasi .....	30
4.2.2 <i>Stopword</i> .....	30
4.2.3 <i>Tf-Idf</i> .....	30
4.2.4 Pencarian .....	31
4.2.4.1 Kendala yang Dihadapi.....	32
4.2.4.2 Solusi yang Ditempuh.....	33
4.3 Pengujian Sistem dan Evaluasi .....	34
4.3.1 Presisi dan <i>Recall</i> .....	36
4.3.2 Perbandingan Hasil Temuan Berdasarkan Jumlah dan Kesamaan Ayat .....	49
BAB V KESIMPULAN DAN SARAN.....	53
5.1 Kesimpulan .....	53
5.2 Saran .....	54
Daftar Pustaka .....	55

## DAFTAR TABEL

Tabel 2.1 Tabel Perhitungan <i>Tf-Idf</i> .....	17
Tabel 2.2 Tabel Perhitungan Model Ruang Vector.....	18
Tabel 2.3 Tabel hasil perhitungan Model Ruang Vector .....	19
Tabel 4.1 Perbandingan hasil pencarian ayat Program, Alkitab Elektronik, dan Sabda .....	35
Tabel 4.2 Perbandingan hasil temuan dengan kata kunci bapa.....	37
Tabel 4.3 Preisi dan <i>recall</i> untuk kata kunci bapa.....	38
Tabel 4.4 Perbandingan hasil temuan dengan kata kunci doa.....	39
Tabel 4.5 Preisi dan <i>recall</i> untuk kata kunci doa .....	40
Tabel 4.6 Perbandingan hasil temuan dengan kata kunci domba.....	41
Tabel 4.7 Preisi dan <i>recall</i> untuk kata kunci domba .....	42
Tabel 4.8 Perbandingan hasil temuan dengan kata kunci kasih karunia .....	43
Tabel 4.9 Preisi dan <i>recall</i> untuk kata kunci kasih karunia.....	44
Tabel 4.10 Perbandingan hasil temuan dengan kata kunci yesus kristus .....	45
Tabel 4.11 Preisi dan <i>recall</i> untuk kata kunci yesus kristus.....	46
Tabel 4.12 Preisi dan <i>recall</i> untuk kata kunci kristen .....	49
Tabel 4.13 Preisi dan <i>recall</i> untuk kata kunci naga.....	50
Tabel 4.14 Preisi dan <i>recall</i> untuk kata kunci sundal.....	51
Tabel 4.15 Preisi dan <i>recall</i> untuk kata kunci merpati.....	51

## DAFTAR GAMBAR

Gambar 2.1 Outline Sistem Temu Kembali Informasi.....	8
Gambar 2.2 Representasi dokumen dan query pada ruang vector.....	10
Gambar 2.3 Representasi matriks kata-dokumen.....	11
Gambar 3.1 Tampilan Awal Sistem.....	20
Gambar 3.2 Tampilan Output.....	21
Gambar 3.3.1 Flowchart Program Utama.....	22
Gambar 3.3.2 Flowchart Pra Proses.....	23
Gambar 3.3.1 Flowchart Pembobotan.....	24
Gambar 3.3.1 Flowchart Pemrosesan kata kunci.....	25
Gambar 3.3.1 Flowchart Pencarian.....	26
Gambar 4.1 Form Menu Utama.....	27
Gambar 4.2 Form Tentang Program.....	28
Gambar 4.3 Form Pencarian.....	28
Gambar 4.4 Form Bantuan.....	29
Gambar 4.5 Pseudocode Tokenisasi dan Stopword.....	30
Gambar 4.6 Pseudocode Penghitungan <i>Tf-Idf</i> .....	31
Gambar 4.7 Hasil pencarian Program dengan kata kunci doa.....	41
Gambar 4.8 Hasil pencarian Alkitab Elektronik dengan kata kunci doa.....	41

Gambar 4.9 Grafik Presisi dan *Recall* hasil pengujian dengan satu kata kunci..... 47

[Gambar 4.10 Grafik Presisi dan \*Recall\* hasil pengujian dengan dua kata kunci  
\(kasih karunia\)..... 47](#)

[Gambar 4.11 Grafik Presisi dan \*Recall\* hasil pengujian dengan dua kata kunci  
\(yesus kristus\) ..... 48](#)

© UKDW

## BAB 1

### PENDAHULUAN

#### 1.1 Latar Belakang

Dalam kehidupan sehari-hari, Alkitab digunakan sebagai salah satu sumber utama dalam mempelajari agama Kristen. Banyak orang Kristen membiasakan diri membaca dan merenungkan firman Allah yang tertulis didalam Alkitab. Akan tetapi, orang Kristen sering kali kesulitan untuk menemukan ayat-ayat yang dibutuhkan dalam waktu cepat. Hal ini disebabkan oleh sangat banyaknya ayat – ayat dalam Alkitab sedangkan memori manusia sangat terbatas untuk menghafal seluruh ayat-ayat dalam Alkitab.

Permasalahan pencarian ayat Alkitab tersebut dapat diselesaikan dengan pencarian otomatis via komputer. Dalam teknologi informasi pencarian otomatis disebut Sistem Temu Kembali atau *Information Retrieval (IR)*. Dengan Sistem Temu Kembali sejumlah dokumen yang besar yang sudah tersimpan dalam suatu file teks dapat ditemukan kembali saat dibutuhkan dengan waktu yang lebih cepat dari pada pencarian secara manual.

Untuk itu penelitian ini mencoba membuat pencarian otomatis untuk menjawab permasalahan pencarian ayat Alkitab tersebut. Jika isi Alkitab dapat diubah menjadi sekumpulan dokumen dalam bentuk file teks maka pencarian otomatis bisa dibuat. Dalam sistem pencarian otomatis tersebut, file tersebut menjadi sumber informasi sedangkan kata-kata yang dicari sebagai kata kunci atau *query*. Permasalahan yang timbul adalah bagaimana membuat pencarian otomatis yang menghasilkan ketepatan dalam waktu yang cepat. Oleh sebab itu penelitian ini diharapkan mampu menghasilkan sebuah sistem pencarian otomatis dengan kemampuan pencarian yang tepat dan cepat sesuai dengan kata kunci atau *query* sehingga pengguna dapat menemukan ayat-ayat dalam Alkitab yang sedang dicari.

## 1.2 Perumusan Masalah

Setelah melihat latar belakang masalah maka disusunlah pokok – pokok rumusan masalah. Adapun masalah-masalah yang akan dibahas dalam laporan ini antara lain:

- a. Bagaimana cara mengimplementasikan *Vector Space Model* (VSM) dalam software pencari ayat Alkitab berdasarkan kata kunci atau *query*?
- b. Bagaimana penyimpanan Alkitab sebagai dokumen?

## 1.3 Batasan Masalah

Batasan masalah yang akan dibahas dalam penulisan tugas akhir ini adalah:

- a. Alkitab yang digunakan adalah Alkitab Terjemahan Baru (1974) Bahasa Indonesia khusus Kitab-kitab didalam Perjanjian Baru terbitan Lembaga Alkitab Indonesia.
- b. Display hasil pencarian kata kunci, berupa ayat-ayat Alkitab yang mengandung kata kunci.

## 1.4 Tujuan Penelitian

Tujuan dari pembuatan Tugas Akhir ini adalah :

- a. Menerapkan konsep *Vector Space Model* dalam sistem pencarian otomatis.
- b. Membantu pengguna untuk menemukan ayat-ayat dalam Kitab Perjanjian Baru yang dibutuhkan dengan cepat dan tepat.

## 1.5 Metode Penelitian

Pada pelaksanaan pembuatan skripsi hingga pembuatan aplikasi dan penyusunan laporan, penulis menggunakan metode yaitu:

- a. Studi Literatur

Pada tahap ini dilakukan pengumpulan data yang diperlukan untuk pembuatan sistem pencarian ayat Alkitab otomatis. Data yang diperlukan diperoleh dengan mempelajari buku, literatur dan jurnal

mengenai permasalahan Information Retrieval dan Vector Space Model. Serta *browsing* melalui internet untuk mencari data yang berhubungan dengan penelitian.

- b. Melakukan analisa ketepatan dan waktu pencarian terhadap kata kunci / *query*.

## 1.6 Sistematika Penulisan

Tugas akhir ini disusun dalam laporan yang dibagi secara sistematis menjadi lima bab, adapun ringkasannya sebagai berikut :

**Bab Pertama adalah Pendahuluan**, yang berisi mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian beserta sistematika penulisan.

**Bab Kedua adalah Tinjauan Pustaka**, menjelaskan teori-teori yang menunjang pemahaman terhadap penelitian yaitu mengenai mesin pencari (*Search Engine*) dan *Information Retrieval* model *Vector Space Model (VSM)*.

**Bab Ketiga adalah Perancangan Sistem**, memuat perancangan implementasi *Vector Space Model* dalam sistem pencarian ayat Alkitab.

**Bab Keempat adalah Implementasi dan Analisis Sistem**, implementasi hasil perancangan dan evaluasi sistem pencarian ayat Alkitab.

**Bab Kelima adalah Kesimpulan dan Saran**, berisikan kesimpulan-kesimpulan dan saran-saran untuk pengembangan sistem selanjutnya.



## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

A. Implementasi dari Vector Space Model dalam Program Pencarian Ayat Kitab Perjanjian Baru adalah sebagai berikut :

1. Mencari token semua dokumen Kitab Perjanjian Baru → ada 7487 token.
2. Mencari total dokumen pasal di Kitab Perjanjian Baru → ada 260 pasal.
3. Mencari frekuensi masing-masing token di tiap-tiap dokumen, kemudian di tambahkan datanya ke tabel *tfidf* (database access)

Proses ini sebenarnya terdiri dari 3 bagian :

- a. Proses mencari frekuensi token ke-*i* di dalam dokumen ke-*j*
  - b. Setelah token ke-*i* sudah dicari frekuensi nya di dalam semua dokumen (260 pasal), maka kita mendapatkan *df* dari token ke-*i* (dalam variabel *df*)
  - c. Menambahkan informasi frekuensi token ke-*i* dalam dokumen ke-*j* ke dalam tabel *tfidf* (tercipta 1.946.620 record, hasil kali dari *n* token dan *n* dokumen ( 7487 \* 260 ) )
4. Mencari *tfidf* tiap token ke-*i* di setiap dokumen ke-*j* (dalam variabel *tfidf*), kemudian menuliskannya secara urut berdasarkan token dan dokumennya ke dalam file text.

Proses ini terdiri dari 2 bagian :

- a. Menghitung *tfidf* dengan rumus  $df(i) * fr(id \text{ token ke-}i, \text{ di dok ke-}j)$ .
- b. Menuliskannya ke dalam file text urut token dan nama dokumennya.

Dengan menyimpan hasil perhitungan Tf-Idf kedalam file.txt membuat akses kedalam file.txt lebih cepat. Cara mengambil nilai Tf-Idf adalah meload file.txt tersebut kemudian menaruhnya kedalam sebuah variabel.

5. Berikutnya masuk kedalam pencarian, kata kunci masukan dari pengguna dihitung *Tf-Idf*-nya. Kemudian dihitung *Cosine*-nya dengan

$$\text{rumus cosine} = \frac{\text{sum}(kk * Di)}{(\text{sqrt}(kk) * \text{sqrt}(Di))}$$

**B.** Cara penyimpanan Kitab Perjanjian Baru sebagai dokumen sebagai berikut, disimpan dalam database access, dengan 3 field, yaitu : Kode kitab, pasal, dan isi. Kode kitab berisi id kitab, pasal berisi pasal-pasal yang terdapat dalam Kitab, isi adalah ayat-ayat dalam pasal tersebut.

## 5.2 Saran

Saran pengembangan yang dapat diberikan pada penulisan Tugas Akhir ini antara lain :

1. *Vector Space Model (VSM)* sebaiknya digunakan dalam Sistem Temu Kembali dengan jumlah dokumen yang kecil.



## DAFTAR PUSTAKA

- Desrianti, Gita. (2010). Akurasi dalam Pencarian pada Search Engines. Bandung : Institut Teknologi Bandung.
- Grossman, Frieder, Goharian. (2002). Retrieval Strategies and Vector Space Model Implementation.
- Harlian, Milka, C.H.R. (2007). Penerapan Web-Based Knowledge Management System Untuk Manajemen Pengalaman Dan Logistik Pasca Bencana Alam. Surabaya : Institut Teknik Sepuluh Nopember.
- Herwansyah, Adhit. Aplikasi Pengkategorian Dokumen Dan Pengukuran Tingkat Similaritas Dokumen Menggunakan Kata Kunci Pada Dokumen Penulisan Ilmiah Universitas Gunadarma. Bandung: Universitas Gunadarma.
- Ilyas, Husni. (2009). Sistem Temu-Balik Informasi. Diakses 17 Maret 2010, dari <http://husni.trunojoyo.ac.id>.
- Ilyas, Husni. (2010). Unified Messaging System Information Retrieval & Klasifikasi Teks. Madura : Universitas Bangkalan.
- Mandala, Rila. (2006). Evaluasi Efektifitas Metode Machine-Learning Pada Search-Engine. Bandung : Institut Teknologi Bandung.
- Novian, Agung. (2004). Panduan Microsoft Visual Basic. Yogyakarta : Penerbit Andi.
- Onarlıoğlu, Kaan. (2009). Information Retrieval Systems. Diakses 9 April 2010 dari <http://www.cs.bilkent.edu.tr/~canf/CS533/cs533spr09hw1possolutions.pdf>.
- Ridok, Achmad., & Furqon, M.T. (2009) Pengelompokan Dokumen Berbahasa Indonesia Menggunakan Metode K-NN. Malang : Ilmu Komputer Universitas Brawijaya.
- Suasanti, Kusuma. (2007). Pembangunan Portal Penyedia Berita Multi-Sumber Berbasis Vector Space Model Untuk Penyedia Berita Berbahasa Inggris. Surabaya : Institut Teknologi Sepuluh Nopember.

Zafikri, Atika. (2008). Implementasi Metode Term Frequency Inverse Document Frequency (Tf-Idf) Pada Sistem Temu Kembali Informasi. Medan : Universitas Sumatera Utara.

© UKDW