

**PEMBENTUKAN VEKTOR DOKUMEN DENGAN
MEMANFAATKAN WORDNET DALAM SISTEM
PENCARIAN INFORMASI DOKUMEN**

TUGAS AKHIR



Oleh

**BONITA CHRISTY
22084472**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
TAHUN 2011**

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul :

**Pembentukan Vektor Dokumen dengan Memanfaatkan WordNet dalam Sistem
Pencarian Informasi Dokumen**

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 kesarjanaaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

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

Yogyakarta, 30-11-2011



Bonita Christy

(22084472)



INTISARI

PEMBENTUKAN VEKTOR DOKUMEN DENGAN MEMANFAATKAN WORDNET DALAM SISTEM PENCARIAN INFORMASI DOKUMEN

Perkembangan teknologi informasi dan globalisasi yang terjadi saat ini berdampak pada meningkatnya kebutuhan pemakai terhadap informasi. Informasi banyak disediakan dalam bahasa Inggris, karena merupakan bahasa global yang dapat dimengerti oleh kebanyakan orang. Guna membantu pemakai dalam mendapatkan informasi, maka dibutuhkan sebuah sistem yang dapat membantu pemakai mencari dokumen yang sesuai dengan kebutuhan informasinya. Hasil pencarian berupa daftar dokumen yang telah diurutkan berdasarkan tingkat relevansinya atau biasa disebut *pe-rangking-an*.

Dalam tiap dokumen teks diperlukan pembentukan vektor guna menentukan bobot tiap kata yang terdapat dalam tiap dokumen teks. Pada pencarian informasi umumnya, pembentukan vektor ini menggunakan setiap kata yang terdapat dalam dokumen, sedangkan dalam bahasa seringkali terdapat sinonim kata yang sama maknanya dengan kata kunci pencarian. Dengan penggunaan sinonim dalam pembentukan vektor, diharapkan sistem dapat membantu pemakai dalam melakukan penemuan kembali dan menghasilkan dokumen-dokumen yang relevan dengan kata kunci berdasarkan pemahaman kata dalam dokumen teks.

Sistem yang dibangun telah dapat menyimpan dokumen yang didapat dari situs www.cnn.com secara lokal. Sistem pencarian dengan metode *Vector Space Model* yang pembentukan vektornya menggunakan *database* WordNet dapat mengembalikan seluruh dokumen yang relevan dengan nilai *precision* 0.447 dan mengembalikan jumlah dokumen relevan paling sedikit ketika nilai *precision* 0.578.

HALAMAN PERSETUJUAN REVISI


Judul : Pembentukan Vektor Dokumen dengan Memanfaatkan
WordNet dalam Sistem Pencarian Informasi Dokumen
Nama : Bonita Christy
NIM : 22084472
Mata Kuliah : Tugas Akhir
Kode : TIW276
Semester : Gasal
Tahun Akademik : 2011/2012

Telah Diperiksa dan disetujui
Di Yogyakarta
Pada Tanggal 29 Januari 2012

Dosen Pembimbing I

Dosen Pembimbing II




Budi Susanto, S.Kom, M.T.



Antonius Rachmat, S.Kom, M.Cs

HALAMAN PENGESAHAN

SKRIPSI

PEMBENTUKAN VEKTOR DOKUMEN DENGAN MEMANFAATKAN
WORDNET DALAM SISTEM Pencarian Informasi Dokumen

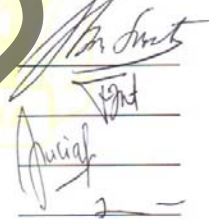
Oleh : Bonita Christy / 22084472

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
19 Desember 2012

Yogyakarta, 9/11/2012
Mengesahkan,

Dewan Penguji :

1. Budi Susanto, S.Kom, M.T
2. Antonius Rachmat, S.Kom, M.Cs
3. Lucia Dwi Krishawati, M.A
4. Sri Suwarno Ir., M.Eng



Dekan Fakultas Teknologi Informasi

Wakil Dekan I Teknik Informatika



(Drs. Wimmie Handiwidjojo, MIT)



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

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kehadira tTuhan Yang Maha Esa yang telah melimpahkan rahmat dan anugerah, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Pembentukan Vektor Dokumen dengan Memanfaatkan Word Net dalam Sistem Pencarian Informasi Dokumen.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa 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. Tuhan Yesus Kristus yang telah memberkati penulis selalu dengan luar biasa, salah satunya dengan menganugerahkan keluarga dan orang-orang sekitar yang selalu mengajarkan untuk bersyukur. Hanya oleh karena berkat-Nya juga penulis bisa ada di titik ini dan menyelesaikan Tugas Akhir.
2. Keluarga tercinta, mami, papi, dan “koko” yang tidak pernah berhenti mendoakan dan mengarahkan.
3. Saudara-saudara yang juga selalu mendoakan dan menginspirasi untuk bias menyelesaikan Tugas Akhir ini;.
4. Bpk. Budi Susanto, S.Kom, M.T. selaku pembimbing I yang telah memberikan bimbingan dan didikan mandiriya dengan sabar dan baik kepada penulis, juga kepada

5. Bpk. Antonius Rachmat, S.Kom, M.Cs. selaku dosen pembimbing II atas bimbingannya, petunjuk, dan masukan yang diberikan selama pengerjaan tugas akhir ini sejak awal hingga akhir, juga kepada
6. Dosen-dosen Universitas Kristen Duta Wacana yang telah membantu memberikan pengarahan dan masukan kepada penulis.
7. Teman-teman “BIG Famz” yang selalu punya acara bersenang-senang, sepadat apapun kegiatan yang ada.
8. Michael Agusta, Havel Cyrus, Sefia Chandra, RikyLiung, Teduhasa, Rico Adrian, yang merupakan teman-teman seperjuangan kuliah selama beberapa semester terakhir, untuk semangat bersama dalam melewati berbagai masalah, dan AndreanWidya Pradipta yang bersama dengan teman semua sama-sama berjuang sampai dengan pengerjaan skripsi.
9. Michelle Amalia, Hernawan, Deddy Wijaya Suliantoro, dan Andrew Siswanto, yang telah banyak membantu dalam masa perkuliahan, baik dukungan maupun pengajaran.

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. Sekali lagi penulis memohon maaf yang sebesar-besarnya. Dan semoga dapat berguna bagi kita semua.

Yogyakarta, 6 Januari 2012

Penulis

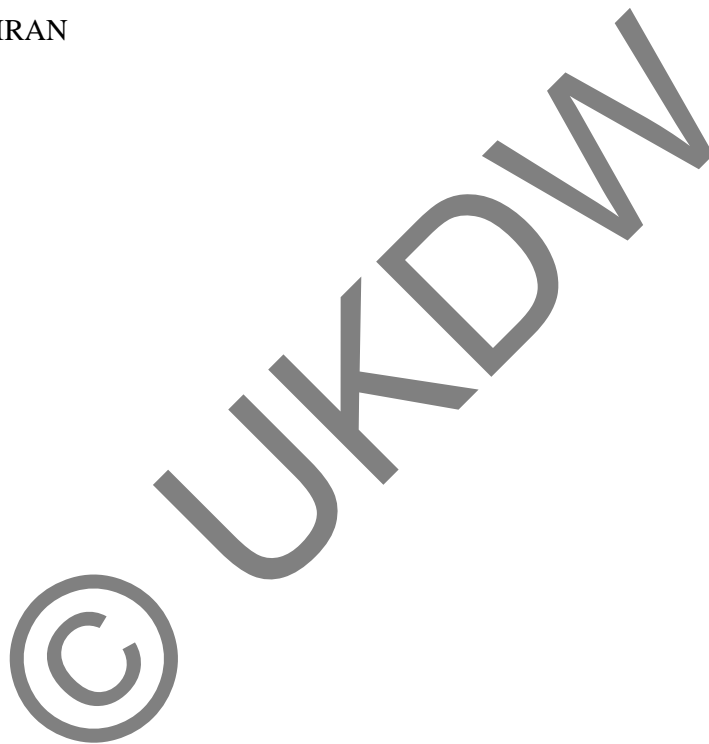
DAFTAR ISI

HALAMAN JUDUL.....	
PERNYATAAN KEASLIAN SKRIPSI.....	ii
INTISARI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
BAB 1 PENDAHULUAN.....	
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah.....	1
1.3 Batasan Masalah.....	2
1.4 Hipotesis.....	2
1.5 Tujuan Penelitian.....	2
1.6 Metode Penelitian.....	2
1.7 Sistematika Penulisan.....	3
BAB 2 TINJAUAN PUSTAKA.....	
2.1 Tinjauan Pustaka.....	5
2.2 Landasan Teori.....	6
2.2.1 WordNet.....	6
2.2.2 <i>Information Retrieval</i>	7
2.2.3 <i>Interpolated Precision-Recall</i>	8
2.2.4 <i>Indexing</i>	10
2.2.5 Algoritma <i>TF-IDF</i>	11
2.2.6 <i>TF-IDF Normalization</i>	12

2.2.7	Metode <i>Vector Space Model</i>	14
BAB 3 ANALISIS DAN PERANCANGAN SISTEM		
3.1	Kebutuhan <i>Hardware</i> dan <i>Software</i>	17
3.2	Spesifikasi Sistem	18
3.3	Arsitektur Sistem	19
3.4	Diagram <i>Use Case</i>	20
3.5	Algoritma dan <i>Flowchart</i>	20
3.5.1	Input dan Data Pelatihan	21
3.5.2	Pencarian dengan <i>Vector Space Model</i>	25
3.6	Kamus Data	26
3.6.1	Tabel Dokumen	26
3.6.2	Tabel Token	27
3.6.3	Tabel Vektor	27
3.6.4	Tabel <i>Stoplist</i>	28
3.6.5	Tabel Sinonim	28
3.6.6	Tabel <i>User</i>	29
3.6.7	Tabel <i>Word</i>	29
3.6.8	Tabel <i>Synset</i>	30
3.6.9	Tabel <i>Lexname</i>	30
3.6.10	Tabel <i>Lexrel</i>	30
3.6.11	Tabel <i>Reltype</i>	31
3.7	Diagram Skema	31
3.8	Rancangan Antarmuka Sistem	32
3.8.1	Rancangan Antarmuka Sistem Utama	32
3.8.2	Rancangan Antarmuka Sistem Pencarian	33

3.9	Rancangan Pengujian Sistem	33
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....		
4.1	Implementasi Sistem	36
4.1.1	Halaman <i>Login</i>	36
4.1.2	Halaman Awal.....	36
4.1.2.1	Fitur <i>Logout</i>	37
4.1.3	Halaman Daftar Dokumen	37
4.1.3.1	Fitur <i>Upload</i> - Admin	37
4.1.3.2	Fitur Lihat Dokumen.....	37
4.1.3.3	Fitur Hitung <i>TF-IDF</i> - Admin.....	38
4.1.3.2.1	Proses <i>Presprocessing</i>	38
4.1.3.2.2	Proses Hitung <i>TF-IDF</i>	41
4.1.4	Halaman Pencarian.....	42
4.1.4.1	Fitur Lihat Sinonim.....	45
4.2	Analisis Sistem.....	46
4.2.1	Pengujian Sistem Pencarian dengan <i>Vector Space Model</i>	46
4.2.1.1	Pengujian Pertama.....	49
4.2.1.2	Pengujian Kedua	50
4.2.1.3	Pengujian Ketiga.....	51
4.2.1.4	Pengujian Keempat	52
4.2.1.5	Pengujian Kelima.....	52
4.2.1.6	Pengujian Keenam	53
4.2.1.7	Pengujian Ketujuh.....	54
4.2.1.8	Pengujian Kedelapan.....	56
4.2.1.9	Pengujian Kesembilan.....	57

4.2.1.10 Pengujian Kesepuluh.....	58
4.2.1.11 Rata-rata Hasil Pengujian.....	59
4.2.1.12 Grafik Interpolasi.....	60
BAB 5 KESIMPULAN DAN SARAN	
5.1 Kesimpulan	61
5.2 Saran.....	61
DAFTAR PUSTAKA	62
LAMPIRAN	



DAFTAR TABEL

Tabel 2.1	ContohHasilPerhitungan <i>TF-IDF</i>	12
Tabel 2.2	ContohHasilPerhitunganNormalisasi <i>TF-IDF</i>	13
Tabel 2.3	ContohHasilPerhitungan <i>Vector Space Model</i>	16
Tabel 3.1	TabelDokumen.....	27
Tabel 3.2	Tabel Token.....	27
Tabel 3.3	TabelVektor.....	28
Tabel 3.4	Tabel <i>Stoplist</i>	28
Tabel 3.5	TabelSinonim.....	29
Tabel 3.6	Tabel <i>Utter</i>	29
Tabel 3.7	Tabel <i>Word</i>	29
Tabel 3.8	TabelSynset.....	30
Tabel 3.9	TabelLexname.....	30
Tabel 3.10	TabelLexrel.....	31
Tabel 3.11	TabelReltype.....	31
Tabel 3.12	TabelMasukanSistem.....	34
Tabel 4.1	TabelHasilEvaluasiSistem.....	46
Tabel 4.2	TabelHasilPengujianPertama.....	50
Tabel 4.3	TabelHasilPengujianKedua.....	50
Tabel 4.4	TabelHasilPengujianKetiga.....	51
Tabel 4.5	TabelHasilPengujianKeempat.....	52
Tabel 4.6	TabelHasilPengujianKelima.....	53
Tabel 4.7	TabelHasilPengujianKeenam.....	54
Tabel 4.8	TabelHasilPengujianKetujuh.....	55

Tabel 4.9 Tabel Hasil Pengujian Kedelapan	56
Tabel 4.10 Tabel Hasil Pengujian Kesembilan	57
Tabel 4.11 Tabel Hasil Pengujian Kese sepuluh	58
Tabel 4.12 Tabel Rata-rata Hasil Pengujian	59

© UKDW

DAFTAR GAMBAR

Gambar 2.1	Kumpulan DokumentanDokumen yang Relevan.....	9
Gambar 2.2	Grafik <i>Precision-Recall</i>	10
Gambar 3.1	ArsitekturSistem.....	19
Gambar 3.2	Diagram <i>Use Case</i>	20
Gambar 3.3	<i>Flowchart</i> Input Data Pelatihan.....	21
Gambar 3.4	<i>Flowchart</i> <i>Tokenizing</i>	22
Gambar 3.5	<i>Flowchart</i> MenentukanSinonim.....	23
Gambar 3.6	<i>Flowchart</i> Perhitungan <i>TF-IDF</i>	24
Gambar 3.7	<i>Flowchart</i> PembentukanVektor.....	25
Gambar 3.8	<i>Flowchart</i> Pencariandengan <i>Vector Space Model</i>	26
Gambar 3.9	Diagram Skema.....	32
Gambar 3.10	RancanganAntarmukaSistemUtama.....	32
Gambar 3.11	RancanganAntarmukaSistemPencarian.....	33
Gambar 4.1	Halaman <i>Login</i>	36
Gambar 4.2	HalamanAwalatauHalamanDaftarDokumen.....	36
Gambar 4.3	HalamanLihat Isi Dokumen.....	38
Gambar 4.4	Halaman Pencarian.....	43
Gambar 4.5	HalamanDaftarSinonim.....	45
Gambar 4.6	GrafikInterpolasi.....	60

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi dan globalisasi yang terjadi saat ini berdampak pada meningkatnya kebutuhan pemakai terhadap informasi. Seiring dengan berkembangnya teknologi informasi yang ada, informasi yang tersedia dalam bentuk dokumen teks pun semakin banyak. Informasi banyak disediakan dalam bahasa Inggris, karena merupakan bahasa global yang dapat dimengerti oleh kebanyakan orang. Guna membantu pemakai dalam mendapatkan informasi, maka dibutuhkan sebuah sistem yang dapat membantu pemakai mencari dokumen yang sesuai dengan kebutuhan informasinya. Hasil pencarian berupa daftar dokumen yang telah diurutkan berdasarkan tingkat relevansinya atau biasa disebut *pe-rangking-an*.

Dalam tiap dokumen teks diperlukan pembentukan vektor guna menentukan bobot tiap kata yang terdapat dalam tiap dokumen teks. Pada pencarian informasi umumnya, pembentukan vektor ini menggunakan setiap kata yang terdapat dalam dokumen, sedangkan dalam bahasa seringkali terdapat sinonim kata yang sama maknanya dengan kata kunci pencarian. Dengan penggunaan sinonim dalam pembentukan vektor, diharapkan sistem dapat membantu pemakai dalam melakukan penemuan kembali dan menghasilkan dokumen-dokumen yang relevan dengan kata kunci berdasarkan pemahaman kata dalam dokumen teks.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah di atas, penulis akan meneliti keakuratan proses temu kembali informasi dengan metode *Vector Space Model* yang dipadukan dengan *database WordNet* dalam pembentukan vektornya. Keakuratan proses dihitung dengan parameter *precision* dan *recall* dari daftar dokumen yang menjadi korpus data.

1.3 Batasan Masalah

Guna menghindari kesalahpahaman dan meluasnya pokok bahasan, maka penulis membatasi perumusan masalah sebagai berikut:

- Data yang digunakan adalah dokumen teks berbahasa Inggris.
- Atribut data yang digunakan sebagai *data training* (data pelatihan) berupa isi dari dokumen teks yang berformat .txt yang dikumpulkan secara manual dari situs berita *CNN* (www.cnn.com)
- Pengolahan teks (menentukan sinonim atau kata yang sama makna) menggunakan *database WordNet*.
- Sistem melakukan pembobotan dengan menggunakan algoritma *TF-IDF* dan perangkaian *Vector Space Model*.
- Relasi yang dipakai dari tabel *WordNet* adalah *also see* dan *derivationally related form*, dengan asumsi kedua relasi tersebut menggambarkan hubungan relasi persamaan makna kata atau sinonim.

1.4 Hipotesis

Penggunaan sinonim atau kata yang sama makna dari tiap kata dalam isi dokumen teks dapat menghasilkan proses temu kembali yang cukup akurat ditinjau dari *recall* dengan menggunakan metode *Vector Space Model*.

1.5 Tujuan Penelitian

Melalui penelitian ini, maka tujuan yang ingin dicapai oleh penulis adalah menerapkan penggunaan *database WordNet* dalam pembentukan vektor proses temu kembali menggunakan metode *Vector Space Model* sehingga dapat menghasilkan pencarian yang cukup akurat dengan nilai *recall* yang cukup tinggi.

1.6 Metode Penelitian

Metode yang digunakan oleh penulis dalam penelitian ini adalah:

- Studi pustaka dan literatur

Studi pustaka dilakukan dengan mencari dan mempelajari sumber-sumber pustaka yang berkaitan dengan pencarian informasi dan penggunaan *database WordNet* dalam aplikasi. Sumber-sumber tersebut berupa buku-buku serta sumber *online* di Internet yang dapat dipercaya.

- Observasi
Observasi pada algoritma dan metode pendukung pencarian informasi teks yang digunakan dalam perancangan sistem.
- Pengujian
Pengujian sistem dilakukan untuk mendapatkan hasil dan kesimpulan dari penelitian ini.

1.7 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini dibagi menjadi beberapa bab, yaitu:

Bab 1 Pendahuluan berfungsi untuk memberikan gambaran umum tentang penelitian. Pendahuluan berisi latar belakang masalah, perumusan masalah, batasan masalah, hipotesis, tujuan penelitian, metode penelitian dan sistematika penulisan laporan.

Bab 2 Tinjauan Pustaka, yang terdiri dari 2 bagian utama, yaitu tinjauan pustaka dan landasan teori. Tinjauan pustaka menguraikan berbagai teori yang didapatkan dari berbagai sumber pustaka yang digunakan dalam penelitian. Sedangkan landasan teori memuat penjelasan tentang konsep dan prinsip utama yang diperlukan untuk memecahkan masalah.

Bab 3 Perancangan Sistem, mengenai kebutuhan *hardware* dan *software* minimum bagi penulis dan pengguna, spesifikasi sistem yang dibuat arsitektur sistem, diagram *use case*, algoritma dan *flowchart*, kamus data, diagram skema, rancangan antarmuka sistem dan rancangan pengujian terhadap sistem.

Bab 4 Implementasi dan Analisis Sistem, membahas implementasi dan pengujian sistem yang sudah dibuat pada bab 3, beserta hasil *capture* dan hasil analisis dari sistem yang dibuat.

Bab 5 Kesimpulan dan Saran, berisi tentang kesimpulan dari hasil penelitian yang dilakukan, serta memberikan saran untuk riset agar dapat memberikan hasil yang lebih baik lagi.

© UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan implementasi sistem, maka dapat diperoleh kesimpulan sebagai berikut:

- Sistem pencarian dengan metode *Vector Space Model* yang pembentukan vektornya menggunakan *database* WordNet dapat mengembalikan seluruh dokumen yang relevan dengan nilai *precision* 0.447 dan mengembalikan jumlah dokumen relevan paling sedikit ketika nilai *precision* 0.578.
- Dari keseluruhan rata-rata hasil pengujian didapat nilai *precision* 0.559, yang artinya sistem masih cukup baik dalam mengembalikan dokumen-dokumen yang relevan.

5.2 Saran

Saran untuk pengembangan dan perbaikan sistem ini adalah:

- Perlu adanya perbaikan algoritma yang mempercepat proses pembobotan token dan pencarian sinonim tiap token, karena semakin banyak jumlah data yang diinputkan maka semakin lama pula proses yang dibutuhkan.
- Perlu adanya perbaikan algoritma yang dapat mempercepat proses pencarian dokumen dengan mencari sinonim dari kata kunci yang dimasukkan dan menyamakan sinonim dengan catatan sinonim dari dokumen-dokumen yang diinputkan terlebih dahulu.
- Pengembangan pencarian dengan kata kunci berupa frasa atau dengan penggunaan *Boolean Operator* (misal: *OR*, *AND*).

Daftar Pustaka

About WordNet. (t.thn.). Dipetik May 23, 2011, dari WordNet a Lexical Database for English: <http://wordnet.princeton.edu/wordnet/>

E.Romero, A. (n.d.). *Geometry and Information Retrieval*.

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook*. New York: Cambridge University Press.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. New York: Cambridge University Press.

Nugroho, S. A. (2009). Query Expansion dengan Menggabungkan Metode Ruang Vektor dan Wordnet pada Sistem Information Retrieval. *Jurnal Informatika* .

Salton, G., & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management* , 24(5): 513-523.

