

PERANCANGAN DAN PEMBUATAN APLIKASI PENCARIAN INFORMASI BEASISWA DENGAN MENGGUNAKAN COSINE SIMILARITY

Nama : Andry Kurniawan
NRP : 08.04.1.1.1.00126
Jurusan : Teknik Informatika
Dosen Pembimbing : Firdaus Solihin, S.Kom., M.Kom.
Fika Hastarita Rachman, S.T., M.Eng.

ABSTRAK

Banyak informasi beasiswa yang ada di internet, di satu sisi memunculkan kesulitan bagi para pencari beasiswa untuk mendapatkan informasi beasiswa. Konsep sistem pencarian berbasis *information retrieval* (IR), dapat digunakan untuk proses mencari informasi beasiswa melalui media internet. Perancangan dan pembuatan IR pada penelitian ini menggunakan konsep *vector space model* (VSM). Untuk pengumpulan data informasi beasiswa menggunakan *Web Crawler*, hasil dari proses crawling akan disimpan ke dalam database. Aplikasi *Web Crawler* yang digunakan adalah *Web Crawler Vietspider*. Kesamaan data beasiswa yang akan diintegrasikan oleh cosine similarity. Dengan menggunakan *cosine similarity*, untuk membentuk data beasiswa yang relevan satu sama lain yang dibutuhkan oleh *user* berdasarkan *query* yang dimasukkan. Berdasarkan hasil dari pengujian sistem, didapatkan nilai minimal *cosine similarity* paling baik adalah *threshold* 0,1. Dengan rata-rata presentase nilai *f-measure* atau tingkat efisiensi kinerja sistem ini mencapai **19,9 %**.

Kata kunci: Beasiswa, *Information Retrieval*, *Vector Space Model*, *Web Crawler*, *Cosine Similarity*.

Halaman ini sengaja dikosongkan

KATA PENGANTAR



Puji syukur kehadiran Allah SWT yang telah memberikan Rahmat dan Hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir ini yang berjudul “ **Perancangan dan Pembuatan Aplikasi Pencarian Informasi Beasiswa dengan menggunakan Cosine Similarity**”.

Disusun dalam rangka memenuhi salah satu tugas akademik untuk menyelesaikan studi di Fakultas Teknik, Universitas Trunojoyo Madura, dan memperoleh gelar kesarjanaan Strata Satu (S-1).

Dalam menyelesaikan tugas akhir ini penulis melaksanakan berdasarkan teori-teori yang didapat dalam bangku perkuliahan, literatur, bimbingan dari dosen pembimbing serta pihak lain yang telah membantu memberikan semangat dan bantuan.

Penulis sadar masih banyak kesalahan dan kekurangan dalam menyelesaikan penyusunan tugas akhir ini, untuk itu penulis mohon maaf dan mengharapkan kritik serta saran dari pembaca. Selain itu juga pembaca dapat mengembangkan tugas akhir ini. Semoga buku ini memberikan manfaat dalam menghasilkan sumber daya manusia yang berkualitas.

Dalam penyusunan tugas akhir ini, penulis mendapatkan bantuan dari berbagai pihak baik secara langsung maupun tidak langsung baik moril maupun material. Untuk itu penulis menyampaikan ucapan terima kasih yang sebesar-besarnya. Utamanya kepada:

1. Allah SWT yang telah mengabulkan semua doa-doa, memberikan kemudahan dan keberuntungan kepada penulis.
2. Keluarga tercinta, Bapak dan Ibu yang tak pernah berhenti mendoakan penulis ini.
3. Ketua Prodi Teknik Informatika Universitas Trunojoyo Madura Ibu Arik Kurniawati, S.Kom.,M.T serta Sekretaris Prodi Teknik Informatika Universitas Trunojoyo Madura Ibu Bain Khusnul Khotimah S.T terima kasih banyak atas bantuannya kepada penulis.
4. Dosen Pembimbing I Bapak Firdaus Solihin, S.Kom.,M.Kom. dan Dosen Pembimbing II sekaligus Dosen Wali, Ibu Fika Hastarita Rachman, S.T.,M.Eng. yang dengan penuh perhatian dan kesabaran

bersedia meluangkan waktunya memberikan bimbingan, petunjuk dan semangat kepada penulis dalam menyusun tugas akhir ini.

5. Rektor Universitas Trunojoyo Madura, Bapak Prof. Dr. Arifin MS,
6. Bapak dan Ibu Dosen Fakultas Teknik Universitas Trunojoyo selaku staf pengajar, yang dengan keikhlasan dan ketulusan hati telah memberikan ilmunya kepada penulis selama masa kuliah.
7. Kawan-kawan UKMFT LPM SAINT yang mengajarku menempa diri dalam berorganisasi.
8. Sahabat serta rekan-rekan sejawat serta seluruh pihak yang tidak dapat penulis sebutkan satu persatu, yang juga ikut memberikan dorongan dan semangat kepada penulis. Terima kasih banyak.

Bangkalan, Juli 2014

Penulis

DAFTAR ISI

	Hal
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iii
ABSTRAK.....	v
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xv
DAFTAR ISTILAH	xvii

BAB I PENDAHULUAN

1.1. Latar Belakang	1
1.2. Perumusan Masalah	2
1.3. Tujuan	3
1.4. Batasan Masalah	3
1.5. Metodologi Penelitian	3
1.6. Sistematika Penulisan.....	4

BAB II TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu	5
2.2. <i>Information Retrieval</i>	5
2.2.1. Pengertian	5
2.2.2. <i>Search Engine</i>	6
2.2.3. <i>Preprocessing</i>	7
2.2.4. <i>Web Crawler</i>	8
2.3. <i>Vector Space Model</i>	9
2.4. <i>Cosine Similarity</i>	11
2.5. Beasiswa	12
2.6. Evaluasi.....	13

BAB III RANCANGAN SISTEM

3.1. Analisa Kebutuhan.....	15
3.1.1. Kebutuhan Perangkat Keras.....	15
3.1.2. Kebutuhan Perangkat Lunak.....	15
3.2. Deskripsi Sistem	16
3.3. Perancangan Sistem	16
3.3.1. <i>Flowchart</i> Sistem.....	19

3.3.1.1. <i>Flowchart Vector Space Model</i>	19
3.3.1.2. <i>Flowchart Admin</i>	20
3.3.1.3. <i>Flowchart User dan Penanganan Query</i>	22
3.3.1.4. <i>Flowchart Preprocessing Sistem</i>	24
3.3.1.5. <i>Flowchart indexing, cosine similarity dan</i> perangkingan	25
3.3.1.6. <i>Flowchart Evaluasi</i>	27
3.3.2. Perancangan Database Sistem.....	29
3.3.2.1. <i>Conceptual Data Model (CDM)</i>	29
3.3.2.2. <i>Physical Data Model (PDM)</i>	30
3.3.3. Struktur Tabel <i>Database</i>	30
3.3.3.1 Tabel Daftar Hasil Crawler.....	31
3.3.3.2 Tabel Simpan.....	31
3.3.3.3 Tabel <i>Beasiswa Relevan</i>	31
3.3.3.4 Tabel <i>History Query</i>	31
3.3.3.5 Tabel <i>Index</i>	32
3.3.3.6 Tabel Login	32
3.3.3.7 Tabel <i>Query</i>	32
3.3.3.8 Tabel Stoplist.....	33
3.3.3.9 Tabel <i>Threshold Cosine</i>	33
3.3.3.10 Tabel Vektor	33
3.3.3.11 Tabel Analisa	33

BAB IV IMPLEMENTASI

4.1. Evaluasi Kerja Perangkat Lunak.....	35
4.1.1. Proses User.....	35
4.1.1.1. <i>Input Query</i>	35
4.1.1.2. Proses <i>Retrieve</i> Beasiswa	36
4.1.2. Proses Admin	37
4.1.2.1. Form Login	37
4.1.2.2. Form Analisa <i>Query</i>	38
4.1.2.3. Form <i>Check</i> Beasiswa Relevan.....	39
4.1.2.4. Form Detail Hasil <i>Retrieve</i> dan Hasil Analisa ...	40
4.1.2.5. Form History Pencarian.....	41
4.1.2.6. Form Hasil Uji Coba Keseluruhan.....	42
4.1.2.7. Form <i>Indexing</i> Pencarian.....	43
4.2. Uji Coba Evaluasi Sistem	44
4.2.1. Skenario Uji Coba Evaluasi Sistem.....	44
4.2.2. Pelaksanaan Skenario Uji Coba	44

4.2.2.1. Uji coba <i>query</i> yang terdiri dari 2 kata	45
4.2.2.2. Uji coba <i>query</i> yang terdiri dari 3 kata	48
4.2.2.3. Uji coba <i>query</i> yang terdiri dari 4 kata	51
4.2.3. Rekapitulasi Dan Analisa Hasil Uji Coba	55
4.2.3.1. Rekapitulasi Hasil Uji Coba	55
4.2.3.2. Analisa Hasil Uji Coba.....	58
 BAB V. PENUTUP	
5.1. Kesimpulan	61
5.2. Saran	61
 DAFTAR PUSTAKA	63
DAFTAR RIWAYAT HIDUP	65

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

	Hal
Gambar 2.1. Proses <i>Information retrieval</i>	6
Gambar 2.2. Desain umum <i>Web Crawler</i>	7
Gambar 2.3. Contoh Model Ruang Vektor dengan Dua Dokumen D1 dan D2, Serta <i>Query</i> Q1.....	10
Gambar 2.4. Contoh Matriks <i>Term-Document</i> untuk <i>Database</i> dengan n Dokumen dan t <i>Term</i>	10
Gambar 3.1. Rancangan umum sistem.....	17
Gambar 3.2. <i>Flowchart Vector Space Model</i>	19
Gambar 3.3. <i>Flowchart admin</i> pada <i>preprocessing</i> dan pembobotan	20
Gambar 3.4. <i>Flowchat User</i> pada proses penanganan <i>query</i>	22
Gambar 3.5. <i>Flowchat</i> Proses <i>Preprocessing</i> sistem	24
Gambar 3.6. <i>Flowchart indexing, cosine similarity</i> dan perangkatan.....	25
Gambar 3.7. <i>Flowchart</i> evaluasi	27
Gambar 3.8. Rancangan CDM <i>database</i> sistem.....	29
Gambar 3.9. Rancangan PDM <i>database</i> sistem	30
Gambar 4.1. Form Input <i>query</i>	35
Gambar 4.2. Proses menampilkan hasil <i>retrieve</i> beasiswa.....	36
Gambar 4.3. <i>Form Login</i>	37
Gambar 4.4. Form analisa <i>query</i>	38
Gambar 4.5. Form <i>check</i> beasiswa relevan	39
Gambar 4.6. Form detail hasil <i>retrieve</i> dan hasil analisa	40
Gambar 4.7. Form history pencarian.....	41
Gambar 4.8. Form hasil uji coba analisa <i>query</i>	42
Gambar 4.9. Form <i>Indexing</i>	43
Gambar 4.10. Grafik uji coba nilai cosine similarity dengan tanpa threshold, 0,1 threshold, 0,2 threshold, 0,3 threshold, dan 0,4 threshold, dari query 2 kata, 3 kata dan 4 kata	58
Gambar 4.11. Grafik nilai rata-rata hasil uji coba evaluasi sistem pada recall dan precision	59
Gambar 4.12. Grafik nilai rata-rata hasil uji coba evaluasi sistem pada f-measure.....	60

Halaman ini sengaja dikosongkan

DAFTAR TABEL

	Hal
Tabel 3.1. Meta	31
Tabel 3.2. Simpan	31
Tabel 3.2. Tb_dok_relevan	31
Tabel 3.3. Tb_history_query	32
Tabel 3.4. Tb_index	32
Tabel 3.5. Tb_login	32
Tabel 3.6. Tb_query	32
Tabel 3.7. Tb_stoplist	33
Tabel 3.8. Cos	33
Tabel 3.9. Tb_vektor	33
Tabel 3.10. Analisa	33
Tabel 4.1. Uji coba tanpa pembatasan nilai <i>cosine similarity</i> dengan query 2 kata	45
Tabel 4.2. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,1 dan query 2 kata	46
Tabel 4.3. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,2 dan query 2 kata	46
Tabel 4.4. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,3 dan query 2 kata	47
Tabel 4.5. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,4 dan query 2 kata	47
Tabel 4.6. Uji coba tanpa pembatasan nilai <i>cosine similarity</i> dengan query 3 kata	48
Tabel 4.7. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,1 dan query 3 kata	49
Tabel 4.8. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,2 dan query 3 kata	50
Tabel 4.9. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,3 dan query 3 kata	50
Tabel 4.10. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,4 dan query 3 kata	51
Tabel 4.11. Uji coba tanpa pembatasan nilai <i>cosine similarity</i> dengan query 4 kata	52
Tabel 4.12. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,1 dan query 4 kata	53
Tabel 4.13. Uji coba pembatasan nilai <i>cosine similarity</i> dengan threshold 0,2 dan query 4 kata	53

Tabel 4.14.	Uji coba pembatasan nilai <i>cosine similarity</i> dengan <i>threshold</i> 0,3 dan <i>query</i> 4 kata.....	54
Tabel 4.15.	Uji coba pembatasan nilai <i>cosine similarity</i> dengan <i>threshold</i> 0,4 dan <i>query</i> 4 kata.....	55
Tabel 4.16.	Rekapitulasi hasil uji coba tanpa pembatasan nilai <i>cosine similarity</i>	55
Tabel 4.17.	Rekapitulasi hasil uji coba pembatasan nilai <i>cosine similarity</i> dengan <i>threshold</i> 0,1	56
Tabel 4.18.	Rekapitulasi hasil uji coba pembatasan nilai <i>cosine similarity</i> dengan <i>threshold</i> 0,2	56
Tabel 4.19.	Rekapitulasi hasil uji coba pembatasan nilai <i>cosine similarity</i> dengan <i>threshold</i> 0,3	57
Tabel 4.20.	Rekapitulasi hasil uji coba pembatasan nilai <i>cosine similarity</i> dengan <i>threshold</i> 0,4	57

DAFTAR ISTILAH

<i>Crawler</i>	Aplikasi yang berfungsi untuk mengambil semua informasi yang diperlukan dengan cara menjelajahi link-link yang ada di halaman situs web kemudian disimpan pada database
<i>user</i>	Orang yang menggunakan aplikasi
<i>Threshold</i>	Nilai batas minimal
<i>query</i>	Kata kunci yang diinputkan oleh <i>user</i>
<i>keyword</i>	Pecahan kata yang ada dalam database
<i>Search engine</i>	Aplikasi mesin pencarian
<i>term</i>	Kata
<i>stoplist</i>	Kata hubung dalam bahasa indonesia
<i>Retrieve</i>	Hasil pencarian
<i>Relevan</i>	Hasil pencarian yang sesuai dengan yang diinginkan oleh <i>user</i>
<i>Flowchart</i>	Diagram alir
<i>User oriented measure</i>	Suatu ukuran berdasarkan yang diharapkan oleh <i>user</i>
Rangking	Urutan dari nilai yang terbesar sampai pada nilai yang terkecil
<i>Preprocessing</i>	Proses pertama kali sebelum melakukan proses yang selanjutnya
<i>Indexing</i>	Pemberian nilai yang menentukan bobot suatu kata

Halaman ini sengaja dikosongkan