

PERANCANGAN DAN PEMBUATAN MESIN PENCARI LOWONGAN PEKERJAAN MENGUNAKAN METODE *COSINE SIMILARITY*

Nama : Denny Setyo Raharjo
NRP : 080411100131
Dosen Pembimbing I : Firdaus Solihin, S.Kom., M.Kom.
Dosen Pembimbing II : Iwan Santosa S.T., M.T.

ABSTRAK

Saat ini data dan informasi lowongan pekerjaan di internet jumlahnya sangat banyak dan terus bertambah dari waktu ke waktu. Untuk mengantisipasi membludaknya informasi lowongan pekerjaan maka berkembanglah teknologi mesin pencari. Mesin pencari dalam penelitian ini menggunakan konsep *information retrieval*. Perancangan dan pembuatan mesin pencari ini memanfaatkan *web crawler* untuk pengumpulan data informasi lowongan pekerjaan. Data informasi lowongan pekerjaan diambil dari internet, hasil dari proses *crawling* akan disimpan ke dalam *database*. Aplikasi *web crawler* yang digunakan adalah *vietspider*. Metode *cosine similarity* digunakan untuk mengukur kemiripan antar dokumen yang terkait dengan *query* yang diberikan oleh *user*. Untuk menggunakan aplikasi ini *user* harus memasukkan *query*, sehingga akan didapatkan hasil pencarian lowongan pekerjaan menurut *query user*. Dalam uji coba pertama *user* memasukkan 3 *query* yang mengandung kata “lowongan kerja”, yaitu “lowongan kerja”, “lowongan kerja bumn”, dan “lowongan kerja bumn pertamina”. Hasil yang dikembalikan ke *user* berjumlah sama yaitu 260 dokumen, dengan nilai dari *cosine similarity* berbeda. Dalam uji coba kedua *user* memasukkan 3 *query* yang berbeda sama sekali, yaitu “lowongan bumn”, “jasa raharja persero”, dan “staff railindo global karya”. Hasil yang dikembalikan ke *user* berjumlah berbeda, yaitu masing masing 246, 72, dan 49. Nilai dari *cosine similarity*-nya juga berbeda.

Kata kunci : Lowongan Pekerjaan, Mesin Pencari, *Information Retrieval*, *Web Crawler*, *Cosine Similarity*

DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR.....	ii
DAFTAR ISI	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL	viii
DAFTAR ISTILAH	ix
DAFTAR LAMPIRAN	x

BAB I PENDAHULUAN

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

BAB II STUDI PUSTAKA

2.1	Penelitian Sebelumnya.....	5
2.2	<i>Information Retrieval</i>	6
	2.2.1 Pengertian	6
	2.2.2 Mesin Pencari.....	6
	2.2.3 <i>Web Crawler</i>	7
	2.2.4 <i>Preprocessing</i>	8
2.3	Metode TF/IDF.....	10
2.4	<i>Cosine Similarity</i>	11

BAB III RANCANGAN SISTEM

3.1	Deskripsi Umum Sistem	13
3.2	Perancangan Sistem	16
	3.2.1 <i>Flowcart Admin</i>	16
	3.2.2 <i>Flowcart User</i> dan Penanganan <i>Query</i>	17
	3.2.3 <i>Flowcart Proses Preprocessing</i>	18
	3.2.4 <i>Flowchart cosine similarity</i> dan perangkingan.....	20
3.3	Struktur Tabel Database	21

BAB IV IMPLEMENTASI

4.1	Lingkungan Perancangan Dan Pembuatan Aplikasi	25
-----	---	----

4.1.1	Spesifikasi Perangkat Keras.....	25
4.1.2	Spesifikasi Perangkat Lunak.....	25
4.2	Evaluasi Kerja Perangkat Lunak	26
4.2.1	Proses User	26
4.2.1.1	Input Query	26
4.2.1.2	Proses Retrieve Lowongan pekerjaan	27
4.2.2	Proses Admin.....	28
4.2.2.1	Crawling Data Menggunakan Vietspider	28
4.2.2.2	Penghilangan Tag XML, Tokenisasi, Stopword Removel Dan Stemming Data Hasil Crawling	35
4.2.2.3	Detail Hasil Perhitungan Cosine Similarity Antara Query Dengan Database Hasil Crawling	37
4.2.2.4	History Query.....	37
4.3	Uji Coba Evaluasi Sistem	38
4.3.1	Skenario Uji Coba Evaluasi Sistem	38
4.3.2	Pelaksanaan Skenario Uji Coba Pertama	38
4.3.2.1	Uji coba query yang terdiri dari 2 kata.....	38
4.3.2.2	Uji coba query yang terdiri dari 3 kata.....	39
4.3.2.3	Uji coba query yang terdiri dari 4 kata.....	40
4.3.3	Pelaksanaan Skenario Uji Coba Kedua.....	41
4.3.3.1	Uji coba query yang terdiri dari 2 kata.....	41
4.3.3.2	Uji coba query yang terdiri dari 3 kata.....	42
4.3.3.3	Uji coba query yang terdiri dari 4 kata.....	43
4.3.4	Rekapitulasi Dan Analisa Hasil Uji Coba.....	44
4.3.4.1	Rekapitulasi Hasil Uji Coba.....	44
4.3.4.2	Analisa Hasil Uji Coba.....	45
BAB V PENUTUP		
5.1	Kesimpulan	47
5.2	Saran	48
DAFTAR PUSTAKA		49
LAMPIRAN		51
DAFTAR RIWAYAT HIDUP		71

DAFTAR GAMBAR

	Hal
Gambar 2.1 Desain Umum Mesin Pencari	7
Gambar 3.1 Gambaran Umum Sistem	14
Gambar 3.2 <i>Flowchart</i> Admin.....	26
Gambar 3.3 <i>Flowchart</i> User dan Penanganan Query	17
Gambar 3.4. <i>Flowchart</i> Proses <i>Preprocessing</i>	18
Gambar 3.5. <i>Flowchart</i> cosine similarity dan perangkingan	20
Gambar 4.1.1 Halaman utama mesin pencari	26
Gambar 4.1.2 <i>Box input user query</i>	27
Gambar 4.2 Hasil proses <i>retrieve</i> lowongan pekerjaan	28
Gambar 4.3.1 Tampilan utama <i>vietspider</i>	29
Gambar 4.3.2 <i>Address bar vietspider</i>	29
Gambar 4.3.3 Pilihan jenis <i>encoding sample crawling</i>	30
Gambar 4.3.4 Pencarian bagian HTML <i>sample crawling</i>	30
Gambar 4.3.5 Pemilihan bagian HTML <i>sample crawling</i>	31
Gambar 4.3.6 Pemilihan kategori <i>sample crawling</i>	31
Gambar 4.3.7 Penulisan parameter <i>sample crawling</i>	32
Gambar 4.3.8 Pemilihan parameter <i>sample crawling</i>	32
Gambar 4.3.9 Menyimpan settingan <i>sample crawling</i>	33
Gambar 4.3.10 <i>Crawling</i> data.....	33
Gambar 4.3.11 Penyiapan <i>Crawling</i> data	34
Gambar 4.3.12 <i>Browse</i> data hasil <i>crawling</i>	34
Gambar 4.3.13 Isi data hasil <i>crawling</i>	35
Gambar 4.4.1. Proses penghilangan tag XML dan <i>preprocessing</i>	36
Gambar 4.4.2 Hasil proses penghilangan tag XML dan <i>preprocessing</i>	36
Gambar 4.5 Detail hasil perhitungan <i>cosine similarity</i>	37
Gambar 4.6 Data <i>query user</i>	38
Gambar 4.7 Hasil skenario uji coba pertama 2 kata	39
Gambar 4.8 Hasil skenario uji coba pertama 3 kata	40

Gambar 4.9 Hasil skenario uji coba pertama 4 kata	41
Gambar 4.10 Hasil skenario uji coba kedua 2 kata	42
Gambar 4.11 Hasil skenario uji coba kedua 3 kata	43
Gambar 4.12 Hasil skenario uji coba kedua 4 kata	44

DAFTAR TABEL

	Hal
Tabel 3.1 Tabel data_mentah.....	21
Tabel 3.2 Tabel data_matang.....	22
Tabel 3.3 Tabel stopword	22
Tabel 3.4 Tabel kamus.....	22
Tabel 3.5 Tabel <i>token</i>	23
Tabel 3.6 Tabel hasil_cs	23
Tabel 3.7 Tabel hasil_cs_detail	23

DAFTAR ISTILAH

<i>Crawler</i>	Aplikasi yang berfungsi untuk mengambil semua informasi yang diperlukan dengan cara menjelajahi <i>link-link</i> yang ada di halaman situs web untuk disimpan pada database
<i>User</i>	Orang yang menggunakan aplikasi
<i>Query</i>	Kata kunci yang diinputkan oleh <i>user</i>
<i>Keyword</i>	Pecahan kata yang ada dalam <i>database</i>
<i>Search engine</i>	Aplikasi mesin pencarian
<i>Term</i>	Kata
<i>Stoplist/Stopword</i>	Kata hubung dalam bahasa indonesia
<i>Retrieve</i>	Hasil pencarian
Relevan	Hasil pencarian yang sesuai dengan yang diinginkan oleh user
<i>Flowchart</i>	Diagram alir
<i>Rangking</i>	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

DAFTAR LAMPIRAN

	Hal
A. Kode Sumber conn_db.php	51
B. Kode Sumber fungsi.php :.....	53
C. Kode Sumber cosine_similarity.php	58
D. Kode Sumber index.php :.....	63