

ABSTRAK

Mesin pencarian telah banyak digunakan di era saat ini untuk membantu *user* dalam mendapatkan informasi. Mesin pencarian membutuhkan *query* untuk menghasilkan informasi yang dibutuhkan oleh *user*. Namun ketika *query* dimasukkan, seringkali masukan tersebut tidak tepat, namun *user* yakin bahwa informasi yang dibutuhkan ada di mesin pencarian tersebut. Pada mesin pencarian saat ini, *user* hanya mendapatkan hasil dari pencarian sesuai dengan yang dimasukkan, meskipun seringkali *query* tersebut kurang spesifik. Pada penelitian ini, menggunakan metode *Relevance feedback* untuk meningkatkan pencarian dokumen dengan memperbaiki *query* awal dengan yang baru, yang lebih spesifik. Proses di dalam pencocokan dokumen dengan *query* menggunakan Algoritma genetika. Terdapat dua metode seleksi yang digunakan, yaitu *roulletewheel selection* dan *tournament selection*. Hasil dari penelitian dengan menggunakan 1000 dokumen tugas akhir ini diperoleh *precision* rata-rata 58%, *recall* rata-rata 89%, dan *f-measure* rata-rata 59% pada penggunaan *roulletewheel selection*. Sedangkan pada penggunaan *tournament selection*, diperoleh *precision* rata-rata 57%, *recall* rata-rata 92%, dan *f-measure* rata-rata 60%. Proses pencarian dengan menggunakan *roulletewheel selection* membutuhkan waktu rata-rata 10.23 detik, dan *tournament selection* 5.97 detik.

Kata Kunci: *Mesin Pencari , Relevance feedback, Roulletewheel Selection, Tournament Selection*

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	iv
KATA PENGANTAR.....	v
HALAMAN PERSEMBAHAN	vi
ABSTRAK	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR KODE PROGRAM	xiii
DAFTAR ISTILAH	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.2.1. Permasalahan.....	3
1.2.2. Usulan Solusi.....	4
1.2.3. Pertanyaan Penelitian	4
1.3 Tujuan dan Manfaat	4
1.3.1. Tujuan.....	4
1.3.2. Manfaat.....	4
1.4 Batasan Masalah.....	4
1.5 Sistematika Laporan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Landasan Teori	6
2.1.1. <i>Information Retrieval</i>	6
2.1.2. <i>Pre-Processing</i>	6
2.1.3. <i>TF-IDF</i>	6
2.1.4. <i>Cosine similarity</i>	7
2.1.5. <i>Relevance Feedback</i>	8
2.1.6. <i>Algoritma Genetika</i>	10
2.2 Penelitian Terkait.....	15
BAB III METODE PENELITIAN	21
3.1 Perancangan Sistem.....	21
3.1.1. <i>Arsitektur Sistem</i>	21
3.1.2. <i>Flowchart Relevance feedback</i>	22

3.1.3.	Flowchart Algoritma Genetika	23
3.1.4.	Flowchart Algoritma Genetika dengan <i>Roulettewheel Selection</i>	26
3.1.5.	Flowchart Algoritma Genetika dengan <i>Tournament Selection</i>	27
3.1.6.	IPO.....	27
3.1.7.	<i>UseCase</i>	28
3.1.8.	CDM	29
3.1.9.	PDM.....	30
3.1.10.	Struktur Tabel Database	31
3.2	Dataset.....	32
3.3	Perhitungan Manual	32
3.4	Tahapan Penelitian	53
3.5	Pengujian.....	54
BAB IV HASIL DAN PEMBAHASAN		56
4.1.	Lingkungan Uji Coba	56
4.2.	Data	56
4.3.	Implementasi User Interface	56
4.3.1.	Halaman awal	56
4.3.2.	Halaman Hasil Pencarian.....	57
4.3.3.	Halaman Feedback Roulettewheel	58
4.3.4.	Halaman Feedback Tournament.....	59
4.4.	Pembangunan Sistem	60
4.4.1.	Pre-Processing	60
4.4.2.	Menghitung TF-IDF	62
4.4.3.	Perhitungan Kemiripan Menggunakan <i>Cosine Similarity</i>	63
4.4.4.	Menghitung Algoritma genetika dengan Roulettewheel selection.....	65
4.4.5.	Menghitung Algoritma genetika dengan Tournament Selection.....	68
4.4.6.	Update generasi	70
4.4.7.	Hasil kemiripan.....	70
4.5.	Pengujian	72
4.5.1.	Proses Pengujian.....	72
4.5.2.	Hasil Pengujian dan Analisa.....	72
4.5.3.	Grafik Hasil Pengujian	82
4.5.4.	Grafik Trend Hasil Pengujian.....	90
4.5.5.	Analisis Hasil Pengujian.....	93
BAB V KESIMPULAN		95

REFERENSI.....	96
LAMPIRAN.....	98

DAFTAR GAMBAR

Gambar 2. 1 Proses Manual Relevance Feedback	8
Gambar 2. 2 Proses Pseudo-Relevance Feedback.....	9
Gambar 2. 3 Istilah dalam Algoritma Genetika	10
Gambar 2. 4 Inisialisasi populasi awal.....	11
Gambar 2. 5 Ilustrasi Roulletewheel selection.....	12
Gambar 2. 6 Ilustrasi Tournament selection	13
Gambar 2. 7 Ilustrasi proses Crossover satu titik.....	14
Gambar 2. 8 Ilustrasi Proses Crossover dua titik	14
Gambar 2. 9 Ilustrasi proses Mutasi.....	15
Gambar 3. 1 Arsitektur Sistem.....	21
Gambar 3. 2 Proses Relevance Feedback	22
Gambar 3. 3 Flowchart Algoritma Genetika.....	23
Gambar 3. 4 Flowchart Algoritma Genetika dengan Roulletewheel	26
Gambar 3. 5 Flowchart Algoritma Genetika dengan Tournament.....	27
Gambar 3. 6 IPO	28
Gambar 3. 7 Usecase.....	29
Gambar 3. 8 CDM.....	30
Gambar 3. 9 PDM	30
Gambar 4.1 Halaman Awal.....	57
Gambar 4 .2 Halaman Hasil Pencarian	58
Gambar 4 .3 Halaman Feedback Roulletewheel	59
Gambar 4 .4 Halaman Feedback Tournament.....	60
Gambar 4 5 grafik precision dengan $P_c=0.5$ dan $P_m=0.02$	83
Gambar 4 6 grafik <i>recall</i> dengan $P_c=0.5$ dan $P_m=0.02$	84
Gambar 4.7 grafik f-measure dengan $P_c=0.5$ dan $P_m=0.02$	84
Gambar 4 8 grafik precision dengan $P_c=0.5$ dan $P_m=0.2$	85
Gambar 4 9 grafik recall dengan $P_c=0.5$ dan $P_m=0.2$	85
Gambar 4 10 grafik f-measure dengan $P_c=0.5$ dan $P_m=0.2$	86
Gambar 4 11 grafik precision dengan $P_c=0.7$ dan $P_m=0.02$	87
Gambar 4 12 grafik recall dengan $P_c=0.7$ dan $P_m=0.02$	87
Gambar 4 13 grafik f-measure dengan $P_c=0.7$ dan $P_m=0.02$	88
Gambar 4 14 grafik precision dengan $P_c=0.7$ dan $P_m=0.2$	89
Gambar 4 15 grafik recall dengan $P_c=0.7$ dan $P_m=0.2$	89
Gambar 4 16 grafik f-measure dengan $P_c=0.7$ dan $P_m=0.2$	90
Gambar 4 17 grafik trend precision	91
Gambar 4 18 grafik trend recall	91
Gambar 4 19 grafik trend f-measure	92
Gambar 4 20 grafik trend waktu	93

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	17
Tabel 3. 1 Pembangkitan Kromosom Populasi Solusi	24
Tabel 3. 2 Contoh pembangkitan populasi awal	24
Tabel 3. 3 Dokumen	31
Tabel 3. 4 Bobot_dok	31
Tabel 3. 5 panjang_vektor_dok	32
Tabel 3. 6 idf_dokumen	32
Tabel 3. 7 Retrieved Document	33
Tabel 3. 8 Populasi solusi kedalam kromosom	34
Tabel 3. 9 Pembangkitan kromosom populasi awal	35
Tabel 3. 10 Pengkodean Kromosom Solusi	38
Tabel 3. 11 nilai fitness	39
Tabel 3. 12 Fitness untuk Rouletewheel Selection	40
Tabel 3. 13 Menghitung Probabilitas	40
Tabel 3. 14 Random angka 0-1	41
Tabel 3. 15 Individu Terpilih	42
Tabel 3. 16 Crossover pada Rouletewheel Selection	43
Tabel 3. 17 Mutasi pada Rouletewheel selection	46
Tabel 3. 18 Populasi baru hasil mutasi	46
Tabel 3. 19 Random individu pada populasi	47
Tabel 3. 20 Crossover pada Tournament Selection	49
Tabel 3. 21 Mutasi Pada Tournament Selection	51
Tabel 3. 22 Populasi baru hasil mutasi	52
Tabel 3. 23 Matriks Recall dan Precision	54
Tabel 4. 1 Lingkungan Uji Coba	56
Tabel 4. 2 Query yang akan dijadikan pengujian	72
Tabel 4. 3 hasil pengujian doumen relevan yang dihasilkan	73
Tabel 4. 4 pengujian dengan $P_c=0.5$ dan $P_m=0.02$	73
Tabel 4. 5 pengujian dengan $P_c=0.5$ dan $P_m=0.2$	75
Tabel 4. 6 pengujian dengan $P_c=0.7$ dan $P_m=0.02$	78
Tabel 4. 7 pengujian dengan $P_c=0.7$ dan $P_m=0.2$	80
Tabel 4. 8 Hasil pengujian waktu pencarian	82

DAFTAR KODE PROGRAM

Kode Program 4. 1 Pre-Processing	61
Kode Program 4. 2 Menghitung TF-IDF	62
Kode Program 4. 3 Menghitung cosine similarity	65
kode program 4. 4 Perhitungan algoritma genetika dengan rouletewheel selection	67
kode program 4. 5 perhitungan algoritma genetika dengan tournament selection	69
kode program 4. 6 Update Generasi.....	70
kode program 4. 7 Hasil Kemiripan.....	71

DAFTAR ISTILAH

<i>Relevance feedback</i>	: Umpan balik dari user yang tujuannya mengetahui apakah data yang terretrieve menurut user sesuai atau relevan dengan query yang dimaksudkan
<i>Information retrieval</i>	: Menemukan kembali informasi yang tersimpan dari berbagai sumber (resources) yang relevan atau koleksi sumber informasi yang dicari atau dibutuhkan
<i>User</i>	: Pengguna
<i>Query</i>	: Kata Kunci
<i>Request</i>	: Permintaan
<i>Relevan</i>	: Sesuai
<i>Performance</i>	: Kinerja
<i>Rekombinasi</i>	: Penggabungan
<i>Term</i>	: Kata
<i>Parent</i>	: individu yang akan dilakukan crossover
<i>Vector</i>	: Sebuah besaran yang memiliki arah dan dapat membentuk sudut
<i>Rouletewheel</i>	: Metode seleksi yang memilih <i>parent</i> berdasarkan nilai kecocokannya (fitness).
<i>Tournament</i>	: Metode seleksi yang memilih <i>parent</i> yang menjalankan beberapa "turnamen" di antara beberapa individu (kromosom) yang dipilih secara acak dari populasi.
<i>MAP</i>	: singkatan dari Mean Average Precision
<i>Evolusi</i>	: Perubahan pada sifat-sifat terwariskan suatu populasi organisme dari satu generasi ke generasi berikutnya.
<i>Manipulasi</i>	: Proses rekayasa dengan melakukan penambahan, atau penghilangan terhadap bagian atau keseluruhan sebuah realitas.
<i>Precision</i>	: Proporsi jumlah dokumen yang ditemukan dan dianggap relevan untuk kebutuhan user

- Recall* : Proporsi jumlah dokumen yang dapat ditemukan-kembali oleh sebuah proses pencarian di sistem Information Retrieval
- F-Measure* : Skor tunggal yang menyeimbangkan antara precision dan recall dalam satu angka.