

**IMPLEMENTASI METODE ROCCHIO'S RELEVANCE
FEEDBACK DALAM PERANGKINGAN DOKUMEN
TEXT**

Tugas Akhir



Oleh :

Indra Pramana

22074221

Program Studi Teknik Informatika Fakultas Teknik

Universitas Kristen Duta Wacana Yogyakarta

Tahun 2011

**IMPLEMENTASI METODE ROCCHIO'S RELEVANCE
FEEDBACK DALAM PERANGKINGAN DOKUMEN TEXT**

Tugas Akhir



**Diajukan kepada Fakultas Teknik Teknologi Informasi
Universitas Kristen Duta Wacana sebagai salah satu syarat
dalam memperoleh gelar Sarjana Komputer**



**Oleh :
Indra Pramana
22074221**

**Program Studi Teknik Informatika Fakultas Teknik
Universitas Kristen Duta Wacana Yogyakarta
Tahun 2011**

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul :

IMPLEMENTASI METODE ROCCHIO'S RELEVANCE FEEDBACK DALAM PERANGKINGAN DOKUMEN TEXT

Yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana komputer pada pendidikan sarjana Program Studi Teknik Informatika, Fakultas Teknik 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 dikemudian hari didapati bahwa skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaaan saya.

Yogyakarta,



(Indra Pramana)

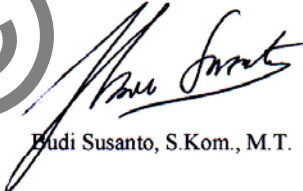
22074221

HALAMAN PERSETUJUAN


Judul : Implementasi Metode Rocchio's Relevance Feedback Dalam
Perangkingan Dokumen Text
Nama : Indra Pramana
NIM : 22074221
Mata Kuliah : Tugas Akhir
Kode : TI2126
Semester : Genap
Tahun akademik : 2010/2011

Telah diperiksa dan disetujui
Di Yogyakarta,
Pada tanggal

Dosen Pembimbing I


Budi Susanto, S.Kom., M.T.

Dosen Pembimbing II


Lucia Dwi Krisnawati C, S.S., M.A.

HALAMAN PENGESAHAN

SKRIPSI

**IMPLEMENTASI METODE ROCCHIO'S RELEVANCE FEEDBACK
DALAM PERANGKINGAN DOKUMEN TEXT**

Oleh Indra Pramana / 22074221

Dipertahankan di depan dewan Penguji Tugas Akhir
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

09 Mei 2011

Yogyakarta, 08/5/2011

Mengesahkan,

Dewan Penguji

1. Budi Susanto, S.Kom., M.T.
2. Lucia Dwi Krisnawati, S.S., M.A.
3. Willy Sudiarto R. S.Kom., M.Cs.
4. Yuan Lukito, S.Kom.







Dekan

Ketua Program Studi



(Drs. Wimmie Handiwidjojo, MIT.)



(Nugroho Agus Haryono, S.Si.,MSi.)

UCAPAN TERIMA KASIH

Terima kasih. Yang pertama dan yang terutama kepada Tuhan Yesus Kristus atas segala hikmat, berkat dan kasih karunia-Nya dalam hidup saya, segala kemuliaan hanya bagi Engkau.

Kepada Papa dan Mama atas segala doa, kasih sayang dan kesabaran yang selalu melimpah bagi Samuel. Terima kasih atas pengalaman hidup yang Papa dan Mama ajarkan bagi Samuel. Tuhan selalu berkati Papa dan Mama.

Kepada koko, cie2 dan seluruh keluarga atas segala doa, kasih sayang dan penghiburan yang kalian berikan. Tuhan berkati kalian semua.

Terima kasih juga kepada dosen pembimbing I, Bapak Budi Susanto. Terima kasih atas segala kesabaran selama membimbing pembuatan Tugas Akhir dan pengetahuan yang telah Bapak berikan. Tuhan selalu Berkati Pak Budsus sekeluarga.

Kepada dosen pembimbing II, Bu Lucia. Terima kasih atas segala kesabaran dan pengetahuan yang Ibu berikan. Tuhan selalu Berkati Ibu Sekeluarga.

Kepada teman-teman yang telah memberikan penghiburan dikala susah. Tertawa itu menular, dan hati yang gembira adalah obat yang manjur. Tuhan berkati kalian semua.

Kepada pihak-pihak yang tidak dapat saya sebutkan satu persatu yang telah membantu baik secara langsung ataupun tidak langsung. Tuhan berkati kalian semua.

Akhir kata saya, selaku penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan maupun yang pernah saya lakukan sewaktu membuat program Tugas Akhir.

Yogyakarta, April 2011

Penulis

INTISARI

IMPLEMENTASI METODE ROCCHIO'S RELEVANCE FEEDBACK DALAM PERANKINGAN DOKUMEN TEXT.

Perkembangan ilmu pengetahuan yang pesat dewasa ini telah mendorong permintaan akan kebutuhan informasi akan ilmu pengetahuan itu sendiri. Akan tetapi semakin luas dan berkembang informasi yang beredar, membuat masyarakat mengalami kesulitan untuk mendapatkan informasi yang dibutuhkannya.

Untuk mempermudah mencari informasi yang dibutuhkan, penulis membangun suatu sistem pencarian dengan menerapkan *relevance feedback*, dengan harapan jika melibatkan pengguna dalam pencarian informasi dapat memberikan hasil pencarian yang lebih baik. Adapun metode *relevance feedback* yang digunakan oleh penulis adalah *metode rocchio's relevance feedback*. Penulis kemudian menguji keakuratan pencarian sistem yang menggunakan *relevance feedback* dalam pencarian dan membandingkannya dengan sistem pencarian yang tidak menggunakan *relevance feedback*.

Dari uji coba sistem pencarian yang menggunakan *rocchio's relevance feedback* ternyata memiliki keakuratan yang lebih tinggi dibandingkan dengan sistem pencarian tanpa menggunakan *relevance feedback*.

Kata kunci : *Rocchio, Relevance Feedback, Information Retrieval*

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN TUGAS AKHIR	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
UCAPAN TERIMA KASIH	iv
INTISARI	v
DAFTAR ISI	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	xi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Hipotesis	3
1.5 Tujuan Penelitian	4
1.6 Metode Penelitian	4
1.6.1 Pra-Pemrosesan	4
1.6.2 Proses	5
1.6.3 Pasca-Pemrosesan	5
1.7 Sistematika Penulisan	5
BAB 2 TINJAUAN PUSTAKA	7
2.1 Tinjauan Pustaka	7
2.2 Landasan Teori	8
2.2.1 Information Retrieval	8
2.2.2 Pembobotan TF-IDF	8
2.2.3 Cosine Similarity	9
2.2.4 Vector Space Model	11

2.2.5	Relevance Feedback	13
2.2.6	Rocchio's Relevance Feedback	14
2.2.7	Stopwords	17
2.2.8	Interpolated Precision-Recall	18
BAB 3 ANALISIS DAN PERANCANGAN SISTEM		21
3.1	Kebutuhan Hardware Dan Software	21
3.1.1	Kebutuhan Hardware	21
3.1.2	Kebutuhan Software	21
3.2	Spesifikasi Sistem	22
3.3	Arsitektur Sistem	23
3.4	Diagram Use Case	24
3.5	Algoritma Dan Flowchart	26
3.5.1	Pra-Pemrosesan	26
3.5.2	Perhitungan TF-IDF	27
3.5.3	Perhitungan Cosine Similarity	28
3.5.4	Proses	28
3.5.5	Perhitungan Rocchio	29
3.5.6	Pasca-Pemrosesan	30
3.6	Kamus Data	31
3.6.1	Tabel File	31
3.6.2	Tabel Token	32
3.6.3	Tabel File_has_token	33
3.6.4	Tabel Stopwords	33
3.6.5	Tabel Keyword	34
3.6.6	Tabel History	34
3.7	Diagram Skema	35
3.8	Rancangan Antarmuka Sistem	36
3.9	Rancangan Evaluasi	37
3.9.1	Rancangan Evaluasi Sistem Pencarian Tanpa Menggunakan Rocchio's Relevance Feedback	37

3.9.2 Rancangan Evaluasi Sistem Pencarian Dengan Pembobotan Rocchio's Relevance Feedback	37
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM	38
4.1 Implementasi Sistem	38
4.1.1 Konfigurasi Awal	38
4.1.2 Antarmuka Sistem	39
4.1.3 Pra-Pemrosesan	41
4.1.4 Proses	43
4.1.5 Pasca-Pemrosesan	44
4.2 Evaluasi Sistem	45
4.2.1 Evaluasi Pencarian Tanpa Menggunakan Rocchio's Relevance Feedback	46
4.2.2 Evaluasi Pencarian Dengan Menggunakan Rocchio's Relevance Feedback	48
4.2.3 Evaluasi Keseluruhan Query	51
4.2.4 Analisa Sistem	56
4.2.5 Kendala Dan Solusi	58
BAB 5 KESIMPULAN DAN SARAN	59
5.1 Kesimpulan	59
5.2 Saran	59

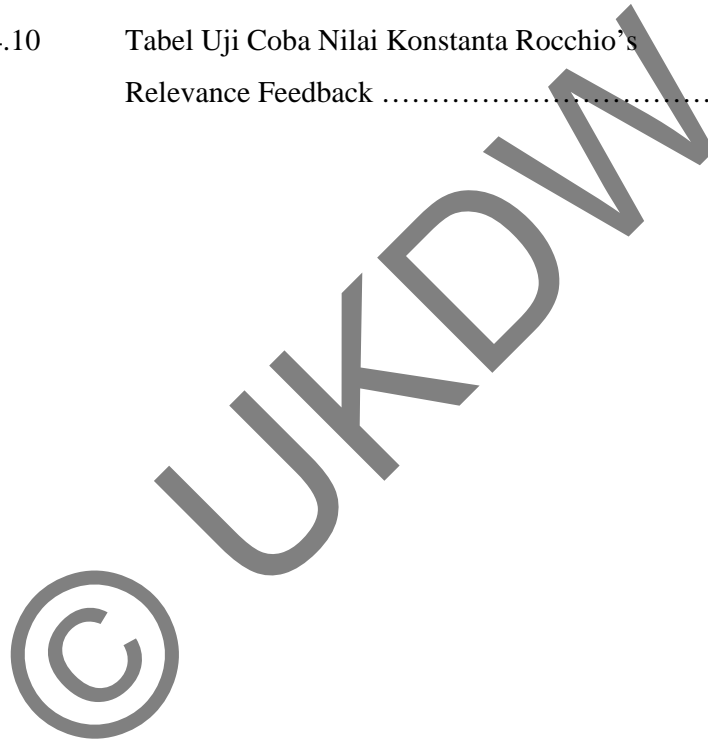
LAMPIRAN

- A. Source Code Program
- B. Hasil Evaluasi
- C. Kartu Konsultasi

DAFTAR TABEL

Tabel 2.1	Indeks Kata	11
Tabel 2.2	Hasil TF-IDF	12
Tabel 2.3	Dokumen Relevan Dan Tidak Relevan	16
Tabel 2.4	Hasil Perhitungan Rocchio	16
Tabel 3.1	Tabel File	32
Tabel 3.2	Tabel Token	32
Tabel 3.3	Tabel File Has Token	33
Tabel 3.4	Tabel Stopwords	34
Tabel 3.5	Tabel Keyword	34
Tabel 3.6	Tabel History	35
Tabel 4.1	Tabel Query Pengujian	46
Tabel 4.2	Tabel Sampel Hasil Perhitungan Non-Interpolated Precision-Recall Tanpa Menggunakan Rocchio's Relevance Feedback	47
Tabel 4.3	Tabel Sampel Hasil Perhitungan Interpolated Precision-Recall Tanpa Menggunakan Rocchio's Relevance Feedback	47
Tabel 4.4	Tabel Sampel Hasil Perhitungan Non-Interpolated Precision-Recall Dengan Menggunakan Rocchio's Relevance Feedback	49
Tabel 4.5	Tabel Sampel Hasil Perhitungan Interpolated Precision-Recall Dengan Menggunakan Rocchio's Relevance Feedback	49
Tabel 4.6	Tabel Keseluruhan Hasil Evaluasi Interpolated Precision-Recall Tanpa Menggunakan Rocchio's Relevance Feedback	52

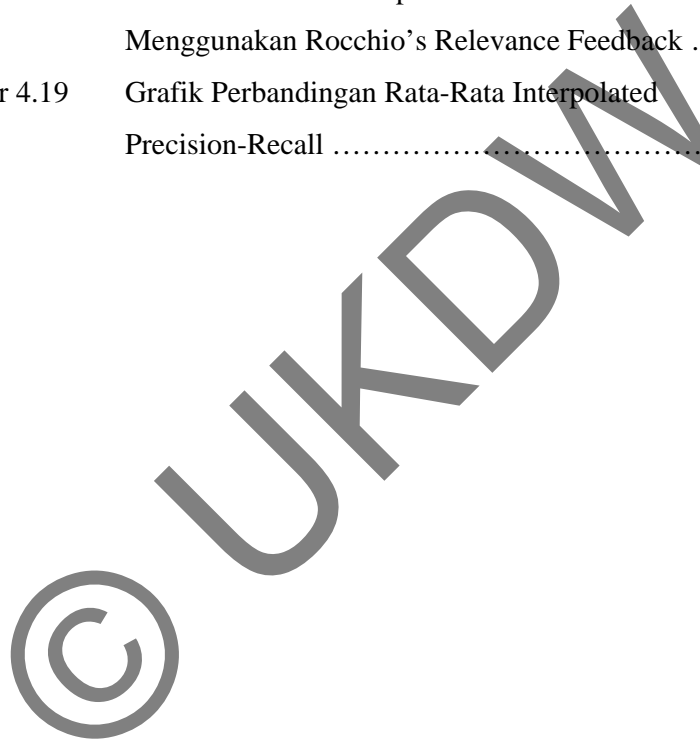
Tabel 4.7	Tabel Rata-Rata Precision Dengan Interpolated Precision-Recall Tanpa Menggunakan Rocchio's Relevance Feedback	52
Tabel 4.8	Tabel Keseluruhan Hasil Evaluasi Interpolated Precision-Recall Dengan Menggunakan Rocchio's Relevance Feedback	54
Tabel 4.9	Tabel Rata-Rata Precision Dengan Interpolated Precision-Recall Dengan Menggunakan Rocchio's Relevance Feedback	54
Tabel 4.10	Tabel Uji Coba Nilai Konstanta Rocchio's Relevance Feedback	57



DAFTAR GAMBAR

Gambar 2.1	Vektor A dan B	10
Gambar 2.2	Relevance Feedback	14
Gambar 2.3	Interpolated Precision-Recall	20
Gambar 3.1	Arsitektur Sistem	23
Gambar 3.2	Diagram Use Case	24
Gambar 3.3	Flowchart Pra-pemrosesan	26
Gambar 3.4	Flowchart Perhitungan TF-IDF	27
Gambar 3.5	Flowchart Perhitungan Cosine Similarity	28
Gambar 3.6	Flowchart Proses	29
Gambar 3.7	Flowchart Perhitungan Rocchio	30
Gambar 3.8	Flowchart Pasca-pemrosesan	31
Gambar 3.9	Skema Database	35
Gambar 3.10	Rancangan Antarmuka Sistem	36
Gambar 4.1	Konfigurasi ODBC Driver	38
Gambar 4.2	Antarmuka Sistem	38
Gambar 4.3	Antarmuka Sistem Dengan Hasil Pencarian	40
Gambar 4.4	File Menu	41
Gambar 4.5	Option Menu	41
Gambar 4.6	Pseudocode Tambah File	42
Gambar 4.7	Pseudocode Insert Token	42
Gambar 4.8	Pseudocode Insert Token Per File	42
Gambar 4.9	Pseudocode Perhitungan Magnitude Query	43
Gambar 4.10	Pseudocode Perhitungan Magnitude Dokumen Dan Dot Product	44
Gambar 4.11	Pseudocode Perhitungan Query Dengan Rocchio	44
Gambar 4.12	Pseudocode Perhitungan Magnitude Dokumen Dengan Rocchio	45
Gambar 4.13	Pseudocode Perhitungan Dot Product Dengan Rocchio ...	45

Gambar 4.14	Grafik Sampel Interpolated Precision-Recall Tanpa Menggunakan Rocchio's Relevance Feedback	48
Gambar 4.15	Grafik Sampel Interpolated Precision-Recall Dengan Menggunakan Rocchio's Relevance Feedback	50
Gambar 4.16	Grafik Sampel Perbandingan Interpolated Precision-Recall	51
Gambar 4.17	Garfik Rata-Rata Interpolated Precision-Recall Tanpa Menggunakan Rocchio's Relevance Feedback	53
Gambar 4.18	Grafik Rata-Rata Interpolated Precision-Recall Dengan Menggunakan Rocchio's Relevance Feedback	55
Gambar 4.19	Grafik Perbandingan Rata-Rata Interpolated Precision-Recall	56



BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan ilmu pengetahuan yang pesat dewasa ini telah mendorong permintaan akan kebutuhan informasi ilmu pengetahuan itu sendiri. Cara pemenuhan kebutuhan akan informasi ini dapat dilakukan dengan beraneka ragam. Mulai dari sekadar membaca koran, majalah, atau jurnal-jurnal tertulis, hingga menggunakan teknologi digital yang terus berkembang.

Akan tetapi semakin luas dan berkembangnya informasi yang beredar, membuat masyarakat mengalami kesulitan untuk mendapatkan informasi yang dibutuhkannya dari media cetak. Lambat laun masyarakat mulai menggunakan teknologi digital untuk memudahkan mereka dalam mencari informasi yang dibutuhkan.

Tetapi hal itu tidak berlangsung lama karena perkembangan informasi berlangsung sangat cepat menyebabkan pencarian akan informasi yang dibutuhkan semakin sulit dan semakin memakan waktu pula. Oleh karena itu salah satu solusi yang digunakan untuk mengatasi masalah ini adalah menggunakan *information retrieval (IR)* untuk memudahkan pencarian informasi agar mendapatkan hasil yang sesuai dengan yang dibutuhkan.

Proses pencarian dan perangkaian dengan menerapkan *relevance feedback* diharapkan dapat menjawab kebutuhan dalam mempermudah pencarian dokumen digital. *Relevance feedback* adalah suatu proses pencarian yang melibatkan *user* dalam pembentukan ulang *query* yang dibutuhkan untuk mencari informasi yang ia butuhkan, dengan memberikan masukan kepada sistem dokumen mana yang dianggap relevan dan dokumen mana yang dianggap tidak relevan. Setelah itu sistem akan melakukan perhitungan ulang berdasarkan

dokumen yang relevan dan yang tidak relevan tersebut, untuk mendapatkan dokumen yang dianggap paling mendekati kebutuhan user. Diharapkan hasil pencarian dari *query* baru tersebut sesuai dengan apa yang *user* inginkan.

Salah satu metode yang ada dalam *relevance feedback* adalah *rocchio's relevance feedback*. *Rocchio's relevance feedback* merupakan suatu metode yang diterapkan pada *vector space model*. Sedangkan untuk pembobotannya metode yang digunakan adalah *tf-idf weighting* dan untuk pengukuran kemiripan antar dokumen digunakan metode *cosine similarity*.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan diatas, maka masalah yang akan diteliti oleh penulis adalah sebagai berikut :

- Bagaimakan cara mengimplementasikan metode *rocchio's relevance feedback* untuk meningkatkan peforma suatu sistem *information retrieval* (IR).
- Apakah dengan menggunakan metode *rocchio's relevance feedback* akan meningkatkan nilai *precision* dan *recall* dalam suatu sistem *information retrieval* (IR).

1.3 Batasan Masalah

Pada penelitian ini, batasan maslaah yang digunakan dalam pembangunan sistem adalah sebagai berikut :

- Dokumen yang digunakan hanya dokumen text dengan format .txt.
- Isi dokumen yang digunakan adalah dokumen yang menggunakan bahasa Indonesia.

- Tidak dilakukan stemming pada setiap *term* yang ada.
- *Stopwords* pada setiap dokumen tidak diindekskan dan tidak disertakan dalam perhitungan.
- *Stopwords* yang digunakan adalah *stopwords* bahasa Indonesia.
- *Query* awal berupa masukan *keyword* dari *user*.
- Jumlah kata pada *keyword* adalah antara satu (1) hingga dua belas (12) kata dan tidak dilakukan *parsing* (tidak menerima operator logika).
- *Output* untuk *query* awal adalah hasil dokumen yang dianggap relevan berdasarkan *keyword* yang diberikan oleh *user*.
- *Feedback* yang diberikan oleh *user* dilakukan dengan menandai dokumen mana saja yang dianggap relevan, dokumen yang tidak ditandai akan dianggap sebagai dokumen yang tidak relevan.
- Pengukuran tingkat kemiripan antar dokumen dengan menggunakan *cosine similarity*.
- Pembobotan untuk *query* awal dihitung dengan menggunakan metode *tf-idf*.
- Pembobotan setelah *user* memberikan *feedback* dilakukan dengan menggunakan *rocchio's relevance feedback*.

1.4 Hipotesis

Sistem Pencarian dengan menggunakan metode *rocchio's relevance feedback* dapat meningkatkan nilai *precision* dan *recall*, sehingga akan memberikan hasil perankingan yang lebih sesuai dengan permintaan *user*.

1.5 Tujuan Penelitian

Tujuan dari penelitian ini untuk mengetahui seberapa besar peningkatan *precision* dan *recall* pada suatu sistem perangkaan dokumen text yang menggunakan metode *rocchio's relevance feedback*.

1.6 Metode Penelitian

Beberapa pendekatan yang digunakan oleh penulis untuk membantu penelitian ini adalah sebagai berikut :

1.6.1 Pra-Pemrosesan

Pada bagian pra-pemrosesan akan dilakukan pengumpulan dokumen text berbahasa Indonesia dengan format .txt untuk koleksi dokumen, yang mana untuk mengetahui apakah dokumen tersebut benar merupakan dokumen .txt akan digunakan aplikasi bantuan, yaitu DROID (*Digital Record Object IDentification*). Setelah itu setiap karakter special dari setiap dokumen akan dihapus untuk kemudian dilakukan proses tokenisasi. Setelah proses tokenisasi selesai, setiap token yang didapat akan diubah menjadi huruf kecil (*lower case*) dan token yang termasuk stopword akan disaring dengan menggunakan list stopwords yang diperoleh dari http://fpmipa.upi.edu/staff/yudi/stop_words_list.txt. Kemudian token yang lolos proses penyaringan akan dihitung bobotnya untuk mempermudah proses perhitungan variable yang nantinya akan dibutuhkan untuk perhitungan selanjutnya.

1.6.2 Proses

Pada bagian ini akan dilakukan perhitungan bobot setiap *term* atau *token* pada masing-masing dokumen untuk memperoleh bobot setiap dokumen dan perhitungan bobot *term* atau *token query* yang diberikan oleh user. Perhitungan

dilakukan dengan menggunakan *vector space model* dan metode *tf-idf weighting*. Kemudian setelah mendapatkan bobot dokumen, akan dilakukan perhitungan untuk menghitung kemiripan suatu dokumen dengan *query* yang diberikan oleh *user* dengan menggunakan metode *cosine similiarity* dan kemudian akan ditampilkan hasil perhitungannya.

1.6.3 Pasca-Pemrosesan

Pada bagian pasca-pemrosesan ini akan dilakukan perhitungan ulang bobot setiap *term* pada *query* berdasarkan *feedback* dokumen relevan dan tidak relevan dari *user*, dengan menggunakan metode *rocchio's relevance feedback*. Setelah didapat bobot yang baru untuk setiap *term* pada *query*, akan dihitung kembali kemiripan suatu dokumen dengan *query* dari *user* dengan menggunakan metode *cosine similiarity* dan kemudian ditampilkan hasilnya.

Setelah didapat hasil perankingan sebelum dan sesudah mendapatkan *feedback* dari *user* akan dilakukan pengujian untuk mengetahui perubahan *precision* dan *recall* pada sistem yang menggunakan *rocchio's relevance feedback*. Pengujian akan dilakukan dengan menghitung nilai *recall* dan *precision* dengan menggunakan *interpolated precision-recall*

1.7 Sistematika Penulisan

Laporan tugas akhir ini secara umum terbagi menjadi lima (5) bab, yaitu :

Bab 1 Pendahuluan, berisi gambaran umum mengenai apa yang diteliti dalam penulisan tugas akhir ini. Pendahuluan ini terdiri dari latar belakang masala, perumusan masalah, batasan masalah, hipotesis, tujuan penelitian, metode penelitian, dan sistematika penulisan laporan tugas akhir.

Bab 2 Tinjauan Pustaka, berisi tinjauan pustaka dan landasan teori yang digunakan dalam penelitian.

Bab 3 Analisis dan Perancangan Sistem, berisi penjelasan mengenai sistem yang akan dibuat , seperti kebutuhan *hardware* dan *software*, spesifikasi sistem, arsitektur sistem, algoritma yang digunakan dalam pembuatan sistem, kamus data, skema *database*, rancangan *user interface*, dan rancangan pengujian sistem.

Bab 4 Implementasi dan Analisis Sistem, berisi pembahasan implementasi dan pengujian sistem yang telah dibuat, hasil implementasi dan analisisnya.

Bab 5 Kesimpulan dan Saran, berisi kesimpulan dari hasil penelitian yang dilakukan dan saran untuk penelitian yang sejenis.



BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan analisis sistem yang menggunakan *rocchio's relevance feedback*, maka diperoleh kesimpulan sebagai berikut :

1. Dalam menggunakan *rocchio's relevance feedback*, nilai vector yang berubah hanyalah vector dari query awal sedangkan nilai vector dari dokumen masih menggunakan perhitungan *tf-idf*.
2. Sistem perangkingan dokumen dengan menggunakan *rocchio's relevance feedback* terbukti mampu memberikan kenaikan nilai *precision* sekitar 3% hingga 9% pada tiap tingkatan *recall*.

5.2 Saran

Beberapa saran yang dianjurkan penulis untuk pengembangan dan perbaikan sistem adalah sebagai berikut :

1. Diperlukan proses *stemming* dan pengecekan sinonim pada proses *ingest* untuk meningkatkan nilai *precision* pada sistem pencarian.
2. Dapat ditambahkan fitur untuk pencarian berupa file text lainnya, seperti *.doc*, *.docx* atau file gambar, seperti *.jpg*, *.png*.
3. Dapat ditambahkan fitur untuk pengelolaan *stopword*.

DAFTAR PUSTAKA

- Garcia, E.(2006). *Cosine Similiarity and Term Weight Tutorial*. Diakses tanggal 17 Januari 2010 dari < <http://www.miislita.com/information-retrieval-tutorial/cosine-similarity-tutorial.html#Cosim>>
- Lee,D.L., Chuang,H., & Seamons,K.(1997). *Document Ranking and the Vector Space-Model*. Diakses tanggal 27 Maret 2011 dari < <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.195&rep=rep1&type=pdf> >
- Mandala,R., & Setiawan,H.(2002) *Peningkatan Performansi Sistem Temu-Kembali Informasi dengan Perluasan Query Secara Otomatis*. Diakses tanggal 27 Maret 2011 dari < <http://belajar.internetsehat.org/pustaka/library-swhw/digital-library/gdl31/IDLN%20III%202002%20Meeting-Paper/Rila%20Mandala-Peningkatan%20Performansi%20Sistem%20Temu%20Kembali.doc> >
- Ramos, J.(2003) *Using TF-IDF to Determine Word Relevance in Document Queries*. Diakses tanggal 27 Maret 2011 dari < <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.1424&rep=rep1&type=pdf> >

Selberg, E.W.(1997). *Information Retrieval Advances Using Relevance Feedback*.

Diakses tanggal 17 Januari 2011 dari <

<http://www.cs.rpi.edu/~chapaa/userSkill/paper/generals.pdf> >

Yates,R.B., & Neto,B.R.,(1999). *Modern Information Retrieval*. England: ACM Press

© UKDW