

**IMPLEMENTASI METODE OKAPI BM11 UNTUK SISTEM TEMU  
KEMBALI INFORMASI**

**TUGAS AKHIR**



Disusun oleh :

**Paulus Rudy Kurnianto**  
**22074216**

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

**IMPLEMENTASI METODE OKAPI BM11 UNTUK SISTEM TEMU  
KEMBALI INFORMASI**

**TUGAS AKHIR**



©  
Diajukan kepada Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
Sebagai salah satu syarat dalam memperoleh gelar  
Sarjana Komputer

Disusun oleh :

**Paulus Rudy Kurnianto**

**22074216**

**Program Studi Teknik Informatika**

**Universitas Kristen Duta Wacana**

**Tahun 2011**

## PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul :

Implementasi Metode Okapi BM11 Pada Sitem Temu Kembali Informasi

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 dikemudian hari didapati bahwa hasil skripsi ini adalah plagiasi atau tiruandari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaaan saya.

Yogyakarta, Juli 2011



( Paulus Rudy Kurnianto )

22074216



HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI METODE OKAPI BM11 UNTUK SISTEM TEMU  
KEMBALI INFORMASI

Oleh : Paulus Rudy Kurnianto / 22 07 4216

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

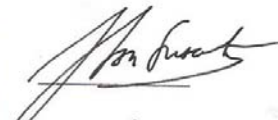
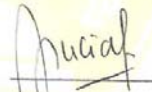
26 Juli 2011

Yogyakarta, 31/8/2011

Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, S.S., M.A.
2. Budi Susanto, S. Kom., M.T.
3. Willy S. R., S.Kom, M.Cs
4. Rosa Delima, M. Kom



Dekan



(Drs. Wimmie Handiwidjojo, MIT)

Ketua Program Studi



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

**HALAMAN PERSETUJUAN**

Judul : Implementasi Metode Okapi BM11 Pada Sitem Temu  
Kembali Informasi

Nama : Paulus Rudy Kurnianto

NIM : 22074216

Mata Kuliah : Tugas Akhir

Kode : TI2126

Semester : Genap

Tahun Akademik : 2010/2011

Telah diperiksa dan disetujui

Di Yogyakarta,

Pada tanggal 18 Juli 2011

Dosen Pembimbing 1

Dosen Pembimbing 2

  
Lucia D. Krisnawati, MA

  
Paulus Rudy Kurnianto

## UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan pada Tuhan Yesus Kristus, yang oleh karena-Nya penulis dapat menjalani masa-masa kuliah dari semester 1 hingga saat ini. Begitu banyak pelajaran yang penulis peroleh, baik dari sisi akademik, maupun dari permasalahan-permasalahan sosial yang penulis alami selama berada di bangku kuliah ini, yang semuanya itu semakin melengkapi pengetahuan dan kemampuan penulis dalam menjalin relasi dengan orang lain. Terima kasih juga penulis ucapkan untuk Papi dan Mami tercinta atas dukungan yang diberikan, doa, dan kasih sayang yang diberikan, juga untuk kakak-kakak tercinta, ko Wawan dan ci Lina yang sudah memberi dukungan dan semangat.

Selain keluarga tercinta, pembuatan tugas akhir ini tidak terlepas dari dukungan pihak Universitas Kristen Duta Wacana. Penulis mengucapkan terima kasih kepada segenap dosen yang membimbing proses belajar selama empat tahun di universitas ini, terkhusus untuk Ibu Lucia Dwi K. dan Bapak Budi Susanto selaku dosen pembimbing yang telah mengarahkan pembuatan tugas akhir ini, serta ko Deddy Wijaya S yang turut membantu mengarahkan dan memberi masukan.

Sahabat dan teman-teman turut menjadi penyemangat dalam menjalani kuliah dan tugas akhir ini. Pertama, terima kasih untuk Eric yang telah menjadi sahabat, teman konsul, sekaligus motivator bagi penulis. Untuk teman-teman *dolan*, Yeye, Erika, Lusi, Anggun, Silvi, Dian, Veve, Peter, Koko, Ian, Andhri, Willi, terima kasih telah menghibur dan menemani jalan-jalan sekedar untuk *refreshing*. Untuk ci Maya, Bagas, Benita, Belinda, Galang, Jihan, Gihon, dan teman-teman GKI Wonogiri lainnya, terima kasih karena selalu menghibur dan menjadi tempat berbagi suka dan duka. Terima kasih untuk teman-teman Duta Voice, kak Sita, mbak Vista yang selama ini menjadi keluarga di Jogja dan bersama-sama melakukan pelayanan.

Masih banyak pihak-pihak yang berperan dalam mendukung penulis menyelesaikan tugas akhir ini, namun tidak dapat disebutkan satu per satu. Penulis hanya dapat mendoakan, semoga pihak-pihak yang telah memberi dukungan dalam bentuk apapun, senantiasa memperoleh berkat, sukacita dan damai sejahtera dari Tuhan Yesus Kristus.

Yogyakarta, Juli 2011

Penulis

© UKDW

## INTISARI

Informasi adalah barang yang teramat penting dalam kegiatan sehari-hari. Baik dalam pendidikan, bisnis, kesehatan, teknologi, dan bidang lainnya, informasi perlu dicari secara cepat, akurat dan mudah. Terlebih lagi dengan berkembangnya teknologi, informasi dapat dibagikan dan diperoleh dari dunia maya atau media penyimpanan pribadi, termasuk informasi yang berupa dokumen teks.

Dalam mencari dokumen yang berisi informasi yang dikehendaki, akan sangat lama dan melelahkan apabila kita harus membuka dan membaca dokumen satu per satu. Untuk mengatasi permasalahan ini, dibutuhkan sistem atau alat yang dapat membantu kita dalam memperoleh dokumen yang mengandung informasi yang kita inginkan secara cepat dan akurat.

Pada penelitian kali ini penulis menggunakan *Information Retrieval (IR)* sebagai solusi. Dengan menggunakan metode *Okapi BM11*, sistem yang dibangun dapat memberi bobot pada dokumen yang mengandung kata-kata yang dicari oleh pengguna, sehingga sistem dapat menyajikan kembali dokumen-dokumen yang kira-kira sesuai/relevan dengan yang diharapkan oleh pengguna.

Dari hasil pengujian terhadap 22 *query*, diperoleh hasil metode *OKAPI BM11* memiliki nilai *precision* yang tinggi dan dapat menampilkan dokumen-dokumen yang relevan terhadap kebutuhan informasi pengguna pada peringkat teratas. Kelemahan dari metode ini ialah setiap token pada *query* seolah-olah menggunakan operator *or*, sehingga semakin panjang *query*, semakin banyak dokumen yang terambil dan akan menurunkan nilai *precision*.



## DAFTAR ISI

HALAMAN SAMPUL DALAM .....	ii
PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
HALAMAN PERSETUJUAN.....	iv
UCAPAN TERIMA KASIH.....	v
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
Bab 1 PENDAHULUAN.....	1
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/Pendekatan .....	3
1.7 Sistematika Penulisan.....	3
Bab 2 LANDASAN TEORI.....	4
2.1 Tinjauan Pustaka .....	4
2.2 Landasan Teori.....	5

1. <i>Information Retrieval</i> (Temu Kembali Informasi).....	5
2. <i>Index Construction</i> .....	6
3. <i>Okapi BM11</i> .....	7
4. <i>Precision dan Recall</i> .....	10
5. <i>Interpolated Precision dan Recall</i> .....	10
<b>Bab 3 ANALISIS DAN PERANCANGAN SISTEM</b> .....	14
3.1 Alat .....	14
3.2 Bahan.....	14
3.3 <i>Use Case</i> .....	15
3.4 Perancangan Masukan.....	18
3.5 Prancangan Keluaran.....	19
3.6 Perancangan Proses .....	19
3.7 Perancangan Basis Data .....	22
3.8 Perancangan Antar Muka .....	23
3.8.1 <i>Form Manipulasi Stop Word</i> .....	24
3.8.2 <i>Form Manipulasi Dokumen</i> .....	25
3.9 Perancangan Evaluasi.....	25
<b>Bab 4 IMPLEMENTASI DAN ANALISIS SISTEM</b>	
4.1 Implementasi Sistem .....	31
4.1.1 Manipulasi Stopword .....	31
4.1.2 Manipulasi Dokumen .....	31

4.1.3 <i>Indexing</i> .....	32
4.1.4 Temu Kembali Informasi .....	34
4.2 Pengujian dan Evaluasi Sistem.....	36
4.2.1 Analisis Sistem.....	36
4.2.2 Evaluasi Sistem .....	39
Bab 5 KESIMPULAN DAN SARAN .....	44
5.1 Kesimpulan.....	44
5.2 Saran.....	44
DAFTAR PUSTAKA .....	45
LAMPIRAN	



© UKDWN

## DAFTAR TABEL

TABEL 2.1 Tabel jumlah dokumen yang berisi tiap token $df_i$ .....	8
TABEL 2.2 Tabel banyaknya sebuah token dalam tiap dokumen.....	8
TABEL 2.3 Tabel idf sebuah token .....	9
TABEL 3.1 <i>Use case text</i> manipulasi <i>stopword</i> .....	15
TABEL 3.2 <i>Use case text</i> manipulasi dokumen .....	16
TABEL 3.3 <i>Use case text</i> proses <i>indexing</i> .....	17
TABEL 3.4 <i>Use case text</i> pencarian` .....	17
TABEL 3.5 Daftar query beserta dokumen yang relevan.....	26
TABEL 4.1 Tabel dokumen, $tf_{id}$ , dan $ D $ untuk <i>query</i> “kartini” .....	36
TABEL 4.2 Hasil keluaran sistem dengan panjang query berbeda .....	37
TABEL 4.3 Hasil temu kembali informasi .....	39
TABEL 4.4 Hasil Temu Kembali Informasi untuk Query “reshuffle kabinet” .....	40
TABEL 4.5 Eleven Point Interpolated Precision .....	41

## DAFTAR GAMBAR

GAMBAR 2.1 Document Retrieval.....	5
GAMBAR 2.2 Grafik Interpolated average precision .....	12
GAMBAR 2.3 Grafik <i>Eleven point average precision</i> .....	13
GAMBAR 3.1 <i>Use case diagram</i> sistem temu kembali informasi.....	15
GAMBAR 3.2 <i>Flowchart</i> Sistem Temu Kembali Informasi.....	20
GAMBAR 3.3 <i>Flowchart</i> Pra Pemrosesan Dokumen dan Query .....	21
GAMBAR 3.4 <i>Flowchart</i> Proses Pembobotan Dokumen .....	21
GAMBAR 3.5 <i>Flowchart</i> Proses Penghitungan Similarity Coefisien.....	22
GAMBAR 3.6 Diagram ER sistem temu kembali informasi.....	22
GAMBAR 3.7 <i>Sitemap</i> sistem temu kembali informasi.....	23
GAMBAR 3.8 Rancangan Antar Muka Halaman Utama.....	23
GAMBAR 3.9 Rancangan Antar Muka Form Manipulasi <i>Stop Word</i> .....	24
GAMBAR 3.10 Rancangan Antar Muka Form Penambahan Dokumen .....	25
GAMBAR 4.1 <i>Pseudocode</i> proses <i>indexing</i> .....	32
GAMBAR 4.2 Algoritma pencocokan <i>stopword</i> dengan token .....	34
GAMBAR 4.3 <i>Pseudocode</i> proses temu kembali informasi .....	35
GAMBAR 4.4 Himpunan token studi dan banding.....	38
GAMBAR 4.5 Grafik precision-recall untuk query “reshuffle kabinet” .....	40
GAMBAR 4.6 Grafik interpolated precision & recall Okapi BM11 .....	43

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Saat ini dokumen teks adalah barang yang sering kita jumpai, baik yang tersimpan di komputer maupun yang berasal dari *Internet*. Dalam melakukan pencarian dokumen teks yang kita inginkan dari sekian banyak dokumen yang ada, akan menjadi sulit jika harus membaca isi dokumen satu demi satu. Oleh karena itu, dibutuhkan sebuah sistem yang mampu menemukan kembali dokumen yang paling relevan.

Sebuah sistem yang disebut *Information retrieval* merupakan sistem yang dapat digunakan untuk menyelesaikan permasalahan tersebut. Dalam *Information Retrieval* dikenal salah satu pendekatan yang dinamakan *probabilistic model*, dimana model ini menitikberatkan pencarian pada hasil perhitungan statistik dan probabilitas dari masing-masing dokumen. *Okapi BM11* merupakan salah satu metode *probabilistic retrieval* yang diperkenalkan oleh Robertson dan Walker(1994) dan akan digunakan dalam penelitian kali ini.

Pada penelitian kali ini, penulis akan mengimplementasikan metode *Okapi BM11* dalam membangun sistem *Information Retrieval*. Dengan mengimplementasikan metode tersebut, untuk menemukan dokumen yang relevan *user* cukup memasukkan *query* kemudian sistem akan menyajikan dokumen-dokumen yang dianggap relevan. Untuk selanjutnya akan dievaluasi kinerja dari sistem ini menggunakan *interpolated precision* dan *recall*.

### 1.2 Perumusan Masalah

Seberapa besar tingkat *precision* dari metode *Okapi BM11*?

### 1.3 Batasan Masalah

Dalam tugas akhir ini, penulis membatasi penelitian dengan beberapa batasan yang didefinisikan di bawah ini:

1. Dokumen teks yang menjadi korpus data adalah *file plain text* yang berformat .txt.
2. Jumlah maksimal *file* .txt yang diteliti adalah 150 *file*.
3. Korpus data yang digunakan adalah berita nasional, regional, dan internasional serta berbahasa Indonesia.
4. *Query* yang diberikan tidak memperhitungkan operator boolean.
5. Tidak dilakukan proses *stemming* untuk setiap *token*.
6. Panjang *query* dibatasi maksimal 5 kata.

### 1.4 Hipotesis

Grafik yang dihasilkan dari perhitungan *interpolated precision* dan *recall* dalam mengevaluasi metode *Okapi BM11* akan mengalami penurunan *precision* tidak lebih dari 40% pada tiap level *recall*.

### 1.5 Tujuan Penelitian

Mengevaluasi kinerja *Okapi BM11* dalam hal pencarian dokumen teks menggunakan *eleven point average precision*.

### 1.6 Metode / Pendekatan

#### 1.6.1 Metode Pengumpulan Data

Untuk memperoleh data berupa dokumen *plain text*, penulis mencari dokumen dari Internet dengan cara mengunduh dari situs-situs web, kemudian dipindah ke dalam sebuah *file* berformat .txt sebagai bahan dalam proses penelitian dan pengujian.

### 1.6.2 Metode Pengembangan Sistem

Dalam pengembangan sistem, metode yang dipakai adalah metode *Okapi BM11* yang termasuk dalam *Probabilistic Retrieval*. Metode *Okapi BM25* digunakan untuk melakukan proses *indexing* dan memperoleh kembali dokumen-dokumen paling relevan berdasarkan *query* yang diberikan oleh *user*.

### 1.6.3 Metode Evaluasi

Dalam mengevaluasi sistem yang akan dibangun, penulis akan menggunakan grafik *interpolated precision*. Akan dilakukan pengamatan terhadap grafik yang dihasilkan untuk 22 *query* yang berbeda. Untuk *query* yang menghasilkan keluaran dokumen dimana jumlah dokumen relevan yang terambil lebih dari 10, akan digambarkan dengan *eleven point average precision*.

## 1.7 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir ini terbagi menjadi lima bab. Bab 1 PENDAHULUAN berisi gambaran umum mengenai penelitian yang akan dilakukan. Bab 2 TINJAUAN PUSTAKA berisi dasar teori yang diperlukan dalam penulisan tugas akhir ini. Bab 3 PERANCANGAN SISTEM membahas tentang bagaimana sistem yang akan dibangun meliputi alur kerja sistem, antar muka sistem, dan perancangan basis data pada sistem. Bab 4 IMPLEMENTASI DAN ANALISIS SISTEM yang akan membahas tentang metode penelitian yang digunakan beserta langkah-langkah serta hasil dari penelitian. Dalam bab ini, diketahui kelebihan dan kekurangan sistem. Bab 5 KESIMPULAN DAN SARAN berisi hal-hal penting yang dapat diambil dari penelitian dan saran pengembangan sistem bagi penelitian berikutnya.



## BAB 5

### KESIMPULAN DAN SARAN

#### 1.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan terhadap metode *OKAPI BM11*, penulis dapat mengambil beberapa kesimpulan :

1. Nilai *similarity coefficient* berbanding lurus terhadap *term frequency*( $t_{fd}$ ) yang terkandung dalam query, dan berbanding terbalik dengan *document length*( $|D|$ ).
2. Semakin panjang *query*, semakin banyak jumlah dokumen yang terambil oleh sistem. Hal ini disebabkan antar token pada query seolah-olah dihubungkan operator *or*.
3. Berdasarkan uji coba terhadap 22 *query* yang dilakukan, *mean average precision* untuk metode *Okapi BM11* yang dihasilkan adalah sebesar 99,279%.

#### 1.2 Saran

Setelah melakukan penelitian ini, penulis memiliki beberapa saran untuk mengembangkan sistem ini, antara lain :

1. Penggunaan operator *boolean* pada query pencarian
2. Menyediakan opsi *advance search*.
3. Dapat melakukan pencarian untuk dokumen berformat .doc dan .pdf.

## DAFTAR PUSTAKA

- Asian, J.(2007). *Effective Techniques for Indonesian Text Retrieval*. RMIT University, Melbourne.
- Clement, D.(2003).*Information Retrieval The Ets Model*. The University of New Brunswick.
- Grossman, D.A., & Frieder, O.(2004). *Information Retrieval, Algorithms and Heuristic*. 2nd Edition. Springer.
- LUK, Robert W.P.(2002). *Different Retrieval Models and Hybrid Term Indexing*.
- Manning, C. D.(2008). *Introduction to Information Retrieval*. Cambridge University Press, New York.
- Robertson, S.E. dan Walker, S.(1994).*Some Simple Effective Approximations To The 2-Poisson Model For Probabilistic Weighted Retrieval* . Proceedings of ACM SIGIR '94 R&D in IR, 1994, pp. 345-354.
- Chi, S.Y. , Hsiao, C. L., Chien, L. F.(2005). *A Practical Passage-based Approach for Chinese Document Retrieval*.

