

**METODE TEMU KEMBALI DALAM PEMBANGUNAN
SITUS PENCARIAN BUKU ELEKTRONIK**

Skripsi



Oleh

Danang Suko Laksono

22064179

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2013

**PENERAPAN *EXTENDED BOOLEAN MODEL P-NORM* SEBAGAI
METODE TEMU KEMBALI DALAM PEMBANGUNAN
SITUS PENCARIAN BUKU ELEKTRONIK**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi Teknik Informatika
Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar
Sarjana Komputer

Oleh

Danang Suko Laksono

22064179

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2013**

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa tugas akhir dengan judul:

Penerapan *Extended Boolean Model P-Norm* Sebagai Metode Temu Kembali Dalam Pembangunan Situs Pencarian Buku Elektronik

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 kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 13 Agustus 2013



(Danang Suko Laksono)

22064179

HALAMAN PERSETUJUAN

Judul : Penerapan *Extended Boolean Model* sebagai metode temu kembali dalam
pembangunan situs pencarian buku elektronik

Nama : Danang Suko Laksono

NIM : 22064179

Mata Kuliah : Tugas Akhir

Kode :

Semester : Genap

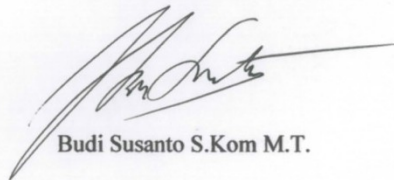
Tahun Akademik : 2012/2013

Telah diperiksa dan disetujui

Di Yogyakarta,

Pada Tanggal 2 Agustus 2013

Dosen Pembimbing I



Budi Susanto S.Kom M.T.

Dosen Pembimbing II



Antonius Rachmat C. S.Kom. M. Cs

HALAMAN PENGESAHAN

PENERAPAN *EXTENDED BOOLEAN MODEL P-NORM* SEBAGAI METODE TEMU KEMBALI DALAM PEMBANGUNAN SITUS PENCARIAN BUKU ELEKTRONIK

Oleh DANANG SUKO LAKSONO/ 22064179

Dipertahankan di depan Dewan Penguji Skripsi
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 13 Agustus 2013

Yogyakarta 22 Agustus 2013

Mengesahkan,

Dewan Penguji :

1. Budi Susanto, S.Kom M.T.
2. Antonius Rachmat C., S.Kom. M.Cs.
3. Yuan Lukito, S.Kom., M.Cs.
4. Theresia Herlina Rochadiani, S.Kom, MT.



Dekan

Drs. Wimmie Handiwidjojo MIT

Ketua Program Studi

Nugroho Agus Haryono, M.Si

UCAPAN TERIMA KASIH

Segala puji dan hormat bagi Dia, **Tuhan Yesus Kristus**, yang telah memberikanku kekuatan, berkat, dan senantiasa membimbingku selama penyusunan Tugas Akhir ini sehingga dapat diselesaikan dengan baik dan tepat waktu.

Penulisan laporan Tugas Akhir ini adalah salah satu bagian kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu, Tugas Akhir ini juga bertujuan melatih penulis menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga bermanfaat bagi penggunanya.

Dalam penyelesaian pembuatan program dan laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, saran dan masukan dari berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu, pada kesempatan ini, dengan segala kerendahan hati penulis ingin mengucapkan terimakasih kepada:

1. **Tuhan Yesus Kristus** yang senantiasa memberiku kekuatan dan berkat. Keajaibannya yang sangat luar biasa telah menolongku dalam segala kesulitan.
2. **Budi Susanto S.Kom M.T**, selaku Dosen Pembimbing I yang telah memberikan bimbingan, petunjuk dan masukan dalam penyelesaian Tugas Akhir ini.
3. **Antonius Rachmat C. S.Kom. M. Cs**, selaku Dosen Pembimbing II yang telah memberikan bimbingan dengan sabar dan baik kepada penulis untuk dapat menyelesaikan Tugas Akhir ini
4. **Bapak dan Mamak** , terimakasih karena kalian yang selalu memberikan aku dukungan dan nasehat dalam segala bentuk moril maupun materil.
5. **Kekasihku “Dian”**, trimakasih atas semuanya, cinta, doa, dukungan dan senantiasa setia mendampingi aku sampai aku dapat menyelesaikan studi.
Love u so much...

6. **Mas Aan, Mbak Ria, Bayu** terima kasih karena telah menyemangati dan tak henti-hentinya mengingatkanku untuk menyelesaikan Tugas Akhir ini.
7. **Hans, Albed, Maichle, Kevin** trimakasih atas segala dukungannya dan bantuannya selama masa kuliah.
8. **Pak Gunardi, Pak Timan, Pak Budi, Mas Edi, Hendra** terima kasih atas doa dan selalu membantu dalam bentuk apapun. Memberi semangat dan mendukung dalam masa pengerjaan Tugas Akhir ini.
9. Rekan-rekan dan pihak-pihak yang tidak dapat penulis sebutkan satu persatu yang secara langsung maupun tidak langsung telah mendukung penyelesaian tugas ini. Terimakasih atas dukungan dan doanya.

Penulis menyadari bahwa laporan dan program Tugas Akhir ini masih jauh dari sempurna, oleh sebab itu penulis sangat mengharapkan kritik dan saran yang membangun dari segenap pembaca sehingga suatu saat penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata penulis ingin meminta maaf bila ada kesalahan baik dalam penyusunan laporan maupun pada saat pembuatan program Tugas Akhir ini. Sekali lagi penulis mohon maaf sebesar-besarnya dan semoga ini dapat berguna bagi kita semua.

Yogyakarta, 13 Agustus 2013

Penulis

DAFTAR ISI

PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	Error! Bookmark not defined.
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Hipotesis.....	2
1.6. Metode Penelitian.....	3
1.7. Sistematika Penulisan.....	3
BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	4
2.1. Tinjauan Pustaka.....	4
2.2. Landasan Teori.....	5
2.2.1. <i>Information Retrieval</i>	5
2.2.2. <i>Index Construction</i>	6
2.2.3. <i>Inverted Index</i>	8
2.2.4. <i>Extended Boolean Model</i>	10
2.2.5. Metode Evaluasi Kinerja.....	15
2.2.6. <i>Electronic Publication (E-Pub)</i>	17
BAB 3 PERANCANGAN SISTEM.....	19
3.1. Spesifikasi Kebutuhan.....	19
3.1.1. Kebutuhan Fungsional.....	19

3.1.2.	Kebutuhan Non-Fungsional	19
3.2.	Arsitektur Sistem	20
3.3.	Perancangan Proses	20
3.4.	Flowchart Alur Sistem Secara Umum	21
3.5.	Perancangan <i>Database</i>	28
3.5.1.	Kamus Data	28
3.5.2.	<i>Entity Diagram Relationship</i>	29
3.6.	Perancangan Antar muka	32
3.6.1.	Rancangan Halaman Utama	29
3.7.	Perancangan Pengujian Sistem	32
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM		37
4.1.	Implementasi Sistem Pencarian Informasi	37
4.2.	Implementasi Sistem Evaluasi Hasil Pencarian	46
4.3.	Analisis Hasil Pencarian	Error! Bookmark not defined.
4.4.	Kelebihan dan Kekurangan Program	62
BAB 5 KESIMPULAN DAN SARAN		63
5.1.	Kesimpulan	63
5.2.	Saran	63
DAFTAR PUSTAKA		64
LAMPIRAN		67

DAFTAR TABEL

Tabel 2.1 Perhitungan <i>idf</i>	13
Tabel 2.2 Perhitungan <i>Tf*idf</i>	13
Tabel 3.1 Tabel Kamus Data.....	28
Tabel 4.1 Tabel data evaluasi.....	49
Tabel 4.2 <i>Eleven-Point Interpolated Precision keyword 1</i>	50
Tabel 4.3 <i>Eleven-Point Interpolated Precision keyword 2</i>	51
Tabel 4.4 <i>Eleven-Point Interpolated Precision keyword 3</i>	52
Tabel 4.5 <i>Eleven-Point Interpolated Precision keyword 4</i>	53
Tabel 4.6 <i>Eleven-Point Interpolated Precision keyword 5</i>	54
Tabel 4.7 <i>Eleven-Point Interpolated Precision keyword 6</i>	55
Tabel 4.8 <i>Eleven-Point Interpolated Precision keyword 7</i>	56
Tabel 4.9 <i>Eleven-Point Interpolated Precision keyword 8</i>	57
Tabel 4.10 <i>Eleven-Point Interpolated Precision keyword 9</i>	58
Tabel 4.11 <i>Eleven-Point Interpolated Precision keyword 10</i>	59
Tabel 4.12 <i>Eleven-Point Interpolated average Precision</i>	60

DAFTAR GAMBAR

Gambar 2.1 <i>Inverted index</i>	9
Gambar 2.2 Proses <i>Inverted index</i>	10
Gambar 2.3 Logika <i>Extended Boolean Model</i>	11
Gambar 2.4 <i>Recall</i> dan <i>Precision</i>	16
Gambar 3.1 Arsitektur Sistem.....	20
Gambar 3.2 <i>Flowchart</i> penyimpanan dokumen.....	21
Gambar 3.3 <i>Flowchart</i> Alur Sistem Proses <i>Text Operation</i>	22
Gambar 3.4 <i>Flowchart</i> Alur Sistem Pencarian Informasi.....	23
Gambar 3.5 <i>Flowchart</i> Alur Sistem Pemrosesan <i>Query</i>	24
Gambar 3.6 <i>Flowchart</i> Alur Sistem Proses <i>IR Similarity Measure</i>	25
Gambar 3.7 <i>Flowchart</i> Alur Sistem Proses Evaluasi.....	26
Gambar 3.8 <i>Flowchart</i> Alur Sistem Proses Evaluasi.....	27
Gambar 3.9 <i>Entity–Relationship Diagrams</i>	29
Gambar 3.10 Rancangan Halaman utama.....	32
Gambar 3.11 Rancangan Halaman <i>Galery.php</i>	33
Gambar 3.12 Rancangan Halaman <i>stopword</i>	33
Gambar 3.13 Rancangan Halaman Evaluasi.....	33
Gambar 3.14 Rancangan Halaman <i>Login</i>	34
Gambar 3.15 Rancangan Halaman <i>Input stopwords</i>	34
Gambar 3.16 Rancangan Halaman <i>Input E-Pub</i>	35
Gambar 3.17 Halaman <i>input</i> evaluasi.....	35
Gambar 4.1 Proses <i>stopwords</i>	37
Gambar 4.2 Memasukan epub kedalam sistem.....	38
Gambar 4.3 Proses membaca epub.....	38
Gambar 4.4 Proses Normalisasi Token e-pub.....	39
Gambar 4.5 Prosedur <i>Add Token</i> ().	40
Gambar 4.6 Prosedur <i>Add Pembagi</i> ().	41
Gambar 4.7 Fungsi validasi <i>query</i>	42
Gambar 4.8 Fungsi Hitung Bobot <i>Query</i>	42

Gambar 4.9 Fungsi Hitung Bobot <i>Query</i> (Lanjutan).	43
Gambar 4.10 Fungsi Perhitungan Jumlah Dokumen.	43
Gambar 4.11 Fungsi Mencari Buku yang mengandung token(Ni).	44
Gambar 4.12 Fungsi mencari Frekuensi (TF).	44
Gambar 4.13 Fungsi perhitungan <i>similarity</i>	45
Gambar 4.14 Hasil pencarian <i>similarity</i>	46
Gambar 4.15 halaman evaluasi.	47
Gambar 4.16 Halaman hasil evaluasi.....	47
Gambar 4.17 Halaman Precision and recall evaluasi.....	48
Gambar 4.18. <i>Eleven-Point Interpolated Precision keyword 1</i>	50
Gambar 4.19 <i>Eleven-Point Interpolated Precision keyword 2</i>	51
Gambar 4.20 <i>Eleven-Point Interpolated Precision keyword 3</i>	52
Gambar 4.21 <i>Eleven-Point Interpolated Precision keyword 4</i>	53
Gambar 4.22 <i>Eleven-Point Interpolated Precision keyword 5</i>	54
Gambar 4.23 <i>Eleven-Point Interpolated Precision keyword 6</i>	55
Gambar 4.24 <i>Eleven-Point Interpolated Precision keyword 7</i>	56
Gambar 4.25 <i>Eleven-Point Interpolated Precision keyword 8</i>	57
Gambar 4.26 <i>Eleven-Point Interpolated Precision keyword 9</i>	58
Gambar 4.27 <i>Eleven-Point Interpolated Precision keyword 10</i>	59
Gambar 4.28 <i>Eleven-Point Interpolated Average Precision</i>	60

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Buku adalah sumber informasi yang sangat berguna, banyak hal yang bisa dipelajari dari buku. Perkembangan teknologi informasi yang terjadi saat ini membuat terobosan dalam kemajuan pembuatan buku. Buku-buku diterbitkan dalam bentuk elektronik dan memanfaatkan internet untuk penyebaran buku elektronik tersebut. *E-pub (electronic publication)* adalah salah satu *file* format penyajian buku elektronik yang terbaru. Penyajian buku elektronik dengan *file* format berupa *e-pub* memiliki kelebihan dibandingkan *file* format buku elektronik yang lain. *E-pub* merupakan kumpulan *file* format *hyper text markup language (html)* sehingga mudah dalam pengidentifikasian *metadata* untuk memproses, menyimpan dan mencari kembali data yang telah dikumpulkan dalam jumlah yang banyak.

Semakin banyak buku diterbitkan dalam bentuk digital maka terjadilah penumpukan jumlah informasi digital yang akhirnya memerlukan pengelolaan informasi digital. Berbagai macam teknologi dibuat untuk mencegah terjadinya ledakan informasi digital, baik di bidang pemrosesan data, penyimpanan data, dan pencarian kembali data digital yang sudah bertumpuk.

Information retrieval (IR) adalah salah satu cabang ilmu pengetahuan yang mempelajari bagaimana mengelola pencarian informasi dalam bentuk digital. IR dapat digunakan untuk membantu dalam pencarian kembali buku-buku elektronik yang disimpan dalam jumlah yang sangat banyak.

Mengingat kemajuan dalam bidang buku elektronik dan *information retrieval* maka penulis berniat membangun sebuah sistem *information retrieval* buku elektronik dalam format *e-pub* dengan memperhatikan *metadata* yang tersimpan sebagai sumber pencarian buku-buku elektronik tersebut.

1.2 Perumusan Masalah

Menerapkan *Extended Boolean Model* sebagai metode temu kembali dalam pembangunan situs pencarian buku elektronik.

1.3 Batasan Masalah

Situs pencarian buku elektronik yang akan dibuat ini memiliki batasan-batasan masalah sebagai berikut :

- a. Situs hanya menggunakan buku elektronik dengan *file* ber format *e-pub* sebagai sumber data
- b. Situs tidak memiliki fasilitas *download*, hal ini dikarenakan untuk melindungi hak cipta dari buku elektronik yang di sediakan, tetapi memiliki fasilitas untuk membaca.
- c. Situs tidak mampu menggunakan perintah pencarian selain boolean moderator

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah menerapkan *Extended Boolean Model* untuk membangun situs pencarian kembali buku elektronik dengan memperhatikan *metadata* yang terdapat dalam buku elektronik dengan format *e-pub* tersebut, sehingga dapat menghasilkan sebuah situs pencarian yang memiliki nilai presisi dan recall yang tinggi.

1.5 Hipotesis

Extended boolean model dapat diterapkan dalam pembangunan situs pencarian kembali buku-buku elektronik dengan memperhatikan *metadata* yang tersimpan sehingga menghasilkan sebuah situs pencarian kembali dengan nilai presisi dan recall yang tinggi.

1.6 Metodologi Penelitian

Pada penelitian ini diterapkan model waterfall dalam metodologi penelitiannya. Adapun tahapan metode penelitiannya adalah sebagai berikut :

1. Analisis format dan metadata buku elektronik dengan format *e-pub* dari beberapa situs standarisasi *e-pub* dan beberapa buku.
2. Analisis penerapan dan kelebihan *extended boolean model* sebagai metode penelitian dari beberapa buku.
3. Perancangan *database* yang akan digunakan dalam pembangunan sistem. Perancangan tabel dan tipe data serta aliran data dalam sistem.
4. Perancangan interface antar muka.
5. Perancangan evaluasi sistem
6. Pengimplemtasian sistem.
7. Pembuatan laporan atau dokumentasi.

1.7 Sistematika Penulisan

Pada laporan tugas akhir ini, penulis membagi laporan menjadi 5 bab. Bab I adalah Pendahuluan yang berisi latar belakang, hipotesis, metodologi penelitian, dan sistematika penulisan. Bab II adalah Tinjauan Pustaka dan Landasan teori, yang menjelaskan teori-teori yang mendukung penulisan tugas akhir ini.

Bab III Analisis dan Perancangan Sistem pada bagian ini membahas analisa dan perancangan sistem meliputi algoritma program, flowchart program dan perancangan input dan output program. Bab IV Implementasi sistem adalah penjelasan bagaimana menggunakan program yang telah dibuat , kesulitan yang dialami pada saat implementasi baik dari segi tampilan maupun proses pembuatan berserta solusi. Bab V Kesimpulan merupakan bab penutup yang berisi kesimpulan dan saran untuk pengembangan program selanjutnya.

BAB 5

KESIMPULAN DAN SARAN

1.1. Kesimpulan

Dari penelitian yang dilakukan oleh penulis, didapatkan hasil pengujian melalui metode evaluasi yang digunakan. Melalui analisis hasil pengujian, didapatkan kesimpulan sebagai berikut :

1. Nilai *average precision* terbesar berada pada level *recall* 0,0 dengan presentase sebesar 70%. Hal ini menunjukkan bahwa buku relevan yang terambil memiliki *similarity* tertinggi pada recall level 0.
2. MAP *Mean average precision* hasil evaluasi sebesar 45%. Hal ini menunjukkan keseluruhan hasil evaluasi bahwa sistem masih bekerja dengan menghasilkan nilai precision yang rendah dibawah 50%.

1.2. Saran

Berikut ini adalah saran yang diperlukan untuk melakukan pengembangan proses pencarian sehingga memberikan hasil yang lebih baik:

1. Korpus data yang digunakan hendaknya memiliki standarisasi yang lebih jelas.
2. Harus ada kejelasan penggunaan *stopwords* dalam kasus penelitian ini adalah bahasa gaul yang sering muncul dalam sebuah novel.
3. Pembacaan novel yang dilakukan dengan baik dan benar akan menentukan hasil evaluasi yang baik.

DAFTAR PUSTAKA

- Baeza-Yates, R., & Riberio-Neto, B. (1999). *Modern Information Retrieval*. Harlow: Addison-Wesley.
- Chaer, A. (1988). *Tata Bahasa Praktis bahasa Indonesia*. Jakarta: Bhratara Karya Aksara.
- Grossman, D.A., & Frieder, F. (2004). *Information Retrieval Algorithm and Heuristics*. Dordrecht: Springer.
- Lee J.H, Kim W.Y, Kim M.H and Lee Y.J (2009) *On the Evaluation of Boolean Operators in the Extended Boolean Retrieval Framework*. Department of Computer Science and Center for Artificial Intelligence Research
- Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Moens, M. (2006). *Information Extraction: Algorithms and Prospects in a Retrieval Context*. Dordrecht: Springer.
- Pohl F, Alistair S, & Zobel J (2010) *Efficient Extended Boolean Retrieval* IEEE Transactions On Knowledge And Data Engineering vol. xx, no. x, month 201x
- Pohl F, Alistair S, & Zobel J (2010) *Extended Boolean Retrieval for Systematic Biomedical Reviews* NICTA Victoria Research Laboratory, Department of Computer Science and Software Engineering The University of Melbourne, Victoria 3010, Australia
- Widodo, J.S (2010). *Pencarian Dokumen Teks Lintas Bahasa Indonesia – Inggris Berbasis Corpus*. Jurnal Teknologi Informasi DINAMIK Volume XV, No.2