

**PERANCANGAN INFORMATION RETRIEVAL MODEL ACAK
BERNOULLI DENGAN PENDEKATAN POISSON**

Tugas Akhir



Oleh:

Aji Pramono

22064163

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Tahun 2012

**PERANCANGAN INFORMATION RETRIEVAL MODEL ACAK
BERNOULLI DENGAN PENDEKATAN POISSON**

Tugas Akhir



**Diajukan kepada Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana**

**Sebagai salah satu syarat dalam memperoleh gelar
Sarjana Komputer**

Disusun oleh:

Aji Pramono

22064163

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Tahun 2012

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

Perancangan Information Retrieval Model Acak Bernoulli dengan Pendekatan Poisson.

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 bagian 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, 28 Juni 2012



AJI PRAMONO
22064163



© UKDWIN

INTISARI

Perancangan Information Retrieval Model Acak Bernoulli dengan Pendekatan Poisson

Divergence From Randomness (DFR) merupakan salah satu model dalam sistem temu kembali informasi yang termasuk dalam model probabilistik. DFR ini terdiri dari instansiasi tiga komponen yaitu pemilihan model dasar, normalisasi pertama dan normalisasi frekuensi term. Salah satu model dalam DFR adalah model acak Bernoulli dengan pendekatan Poisson (metode PL2) dengan normalisasi kedua menggunakan aturan Laplace.

Model acak Bernoulli dengan pendekatan poisson ini bertujuan mencari dokumen relevan berdasarkan inputan pengguna berupa kata kunci. Dokumen yang mengandung kata kunci akan dihitung bobotnya menggunakan metode PL2. Dokumen yang terambil sistem akan ditampilkanurut berdasar nilai relevansi paling tinggi sehingga pengguna dapat memperoleh informasi yang dibutuhkan dengan cepat.

Metode PL2 ini cukup baik dalam mencari dokumen relevan yang dibutuhkan pengguna dalam mencari informasi. Hal ini terlihat dari kinerja sistem yang dihitung menggunakan *interpolated precision* menunjukkan nilai yang tinggi. Dokumen relevan akan ditampilkan pada ranking teratas.

HALAMAN PERSETUJUAN

Judul : Perancangan Information Retrieval Model Acak Bernoulli
dengan Pendekatan Poisson
Nama Mahasiswa : Aji Pramono
NIM : 2206 4163
Mata Kuliah : Tugas Akhir
Kode : TIW276
Semester : Genap
Tahun : 2011 / 2012

Telah diperiksa dan disetujui
di Yogyakarta,
pada tanggal, Mei 2012

Dosen Pembimbing I



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

Dosen Pembimbing II



Drs. R. Gunawan Santosa, M.Si

© UKDW

HALAMAN PENGESAHAN

**PERANCANGAN INFORMATION RETRIEVAL MODEL ACAK BERNOULLI
DENGAN PENDEKATAN POISSON.**

Oleh: AJI PRAMONO / 22064163

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
20 Juni 2012

Yogyakarta, 28 Juni 2012
Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, M.A.
2. Drs. R. Gunawan Santosa, M.Si.
3. Budi Susanto, SKom.,M.T.
4. Aditya Wikan Mahastama, S.Kom



Dekan

(Signature)
(Drs. Wimmie Handiyidjojo, MIT.)

(Signature)
(Signature)
(Signature)

Ketua Program Studi

(Signature)
(Nugroho Agus Haryono, M.Si)

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan anugerah, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Perancangan Information Retrieval Model Acak Bernoulli dengan Pendekatan Poisson dengan baik.

Penulisan laporan ini merupakan kelengkapan dan pemenuhan dari salah satu syarat dalam memperoleh gelar Sarjana Komputer. Selain itu bertujuan melatih mahasiswa untuk dapat menghasilkan suatu karya yang dapat dipertanggungjawabkan secara ilmiah, sehingga dapat bermanfaat bagi penggunanya.

Dalam menyelesaikan 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 dengan segala kerendahan hati, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Lucia Dwi Krisnawati, S.S., M.A. selaku dosen pembimbing I atas bimbingan, petunjuk dan masukan yang diberikan selama pengerjaan tugas akhir ini sejak awal hingga akhir.
2. Bapak Drs. R. Gunawan Santosa, M.Si. selaku dosen pembimbing II yang telah memberikan bimbingannya dengan sabar dan baik kepada penulis.
3. Bapak Ibu dosen Fakultas Teknologi Informasi Universitas Kristen Duta Wacana yang telah membagikan ilmunya kepada penulis sehingga penulis dapat menyelesaikan pendidikan SI-nya.
4. Keluarga tercinta yang memberi dukungan dan semangat.
5. Mas Wendra dan mas Ardhi yang selalu membantu dengan ikhlas dalam pengerjaan program Tugas Akhir ini.
6. Teman – teman TI angkatan '06 yang saling menyemangati dan memberi masukan.
7. Pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik.

Penulis menyadari bahwa program dan laporan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari pembaca sekalian. 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 yang pernah penulis lakukan sewaktu membuat program Tugas Akhir ini. Dan semoga ini dapat berguna bagi kita semua.

Yogyakarta, Mei 2012

Penulis

© UKDWN

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN	v
UCAPAN TERIMA KASIH.....	vi
INTISARI	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB 1 Pendahuluan	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan.....	3
1.5 Metode/Pendekatan	3
1.6 Sistematika Penulisan.....	4
BAB 2 Tinjauan Pustaka.....	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	9
BAB 3 Perancangan Sistem.....	21
3.1 Use Case Diagram	21
3.2 Bahan dan Alat Penelitian	22
3.2.1 Bahan	22
3.2.2 Alat	22
3.3 Perancangan Database	23
3.4 Perancangan Proses	23
3.4.1 Rancangan proses information retrieval	23
3.4.2 Rancangan proses <i>indexing</i>	24

3.4.3 Rancangan proses <i>query</i>	25
3.5 Perancangan Inputan	26
3.5.1 Masukan Dokumen	27
3.5.2 Masukan Query	27
3.5.3 Masukan Stopword	27
3.6 Perancangan Keluaran	27
3.7 Perancangan User Interface	27
3.7.1 Form Masukan Dokumen	28
3.7.2 Rancangan Form Masukan Query dan Retrieval	28
3.7.3 Rancangan form masukan stopwords	29
BAB 4 Implementasi dan Analisis Sistem	30
4.1 Strategi Implementasi	30
4.2 Implementasi Menu Sistem	33
4.2.1 Implementasi Form Retrieval	33
4.2.2 Implementasi Form Input	34
4.2.3 Implementasi Form Stopword	35
4.2.4 Implementasi Form Buka Document	36
4.2.5 Implementasi Form List Document	37
4.3 Proses Keluaran Sistem	37
4.4 Analisis Proses Keluaran Sistem	38
4.5 Kelebihan dan Kekurangan	51
BAB 5 Kesimpulan dan Saran	50
5.1 Kesimpulan	50
5.2 Saran	50

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Tabel	Keterangan	Halaman
2.1	Model dasar Divergence From Randomness	5
2.2	Nilai Parameter c untuk jenis query	7
2.3	Perbandingan MAP dari model DFR	8
2.4	Tabel frekuensi term	16
2.5	Perhitungan interpolasi recall precision	20
4.1	Daftar Dokumen Relevan	39
4.2	Precision dan Recall pengujian 1	40
4.3	Interposisi Precision pengujian 1	41
4.4	Precision dan Recall pengujian 2	41
4.5	Interpolasi Precision pengujian 2	42
4.6	Precision dan Recall pengujian 3	42
4.7	Interpolasi Precision pengujian 3	42
4.8	Precision dan Recall pengujian 4	43
4.9	Interpolasi Precision pengujian 4	43
4.10	Precision dan Recall pengujian 5	43
4.11	Interpolasi Recall Precision pengujian 5	44
4.12	Precision dan Recall pengujian 6	45
4.13	Interpolasi Recall Precision pengujian 6	45
4.14	Precision dan Recall pengujian 7	45
4.15	Interpolasi Precision pengujian 7	46
4.16	Precision dan Recall pengujian 8	46
4.17	Interpolasi Precision pengujian 8	47
4.18	Precision dan Recall pengujian 9	47
4.19	Interpolasi Precision pengujian 9	47
4.20	Precision dan Recall pengujian 10	48
4.21	Interpolasi Precision pengujian 10	48

DAFTAR GAMBAR

Gambar	Keterangan	Halaman
3.1	Use Case Diagram	21
3.2	Rancangan Database	23
3.3	Flowchart proses IR	24
3.4	Flowchart proses indexing	25
3.5	Flowchart proses query	26
3.6	Rancangan form masukan dokumen	28
3.7	Form Retrieval dan masukan query	29
3.10	Form Stopword	29
4.1	Implementasi form retrieval	33
4.2	Implementasi form input data	34
4.3	Implementasi form stopwords	35
4.4	Implementasi form buka dokumen	36
4.5	Implementasi form list dokumen	37
4.6	Hasil keluaran sistem	38
4.7	Hasil retrieval dengan query "bahan bakar minyak"	40
4.8	Implementasi dengan query bolak-balik	44

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan berkembangnya pengetahuan dan teknologi, akses terhadap informasi semakin mudah dan cepat diperoleh oleh banyak orang. Informasi ini dapat berupa dokumen, berita, data keuangan, artikel dan lain-lain. Kemudahan dalam mendapatkan informasi ini menyebabkan banyaknya informasi dan beraneka ragam jenisnya.

Dengan banyaknya informasi yang diperoleh, disadari bahwa masalah utama telah bergeser dari cara mengakses informasi menjadi memilih informasi yang berguna secara selektif. Banyaknya dokumen yang tersimpan di dalam komputer membuat pengguna mengalami kesulitan ketika mencari dokumen yang sesuai kebutuhan. Tidak memungkinkan jika pengguna harus mencari dokumen satu persatu dan membaca setiap dokumen karena hal tersebut memerlukan waktu yang lama dan tidak efisien.

Karena itu, dibutuhkan suatu sistem untuk melakukan pencarian suatu dokumen dalam sekumpulan dokumen untuk memilih dokumen yang sesuai dengan kebutuhan pengguna dengan memasukkan kata kunci (*keyword*). Salah satu sistem untuk pencarian tersebut adalah sistem temu kembali informasi (*information retrieval system-IR*). Dengan sistem ini diharapkan pengguna dapat menemukan dokumen yang dibutuhkan dengan cepat dan akurat.

Information retrieval juga mengalami perkembangan. Berbagai metode ditemukan dan dikembangkan untuk meningkatkan kinerja dari sistem ini. Antara metode satu dengan metode yang lain mempunyai kelebihan dan kelemahan tersendiri. Salah satu metode pencarian dalam temu kembali informasi adalah menggunakan model acak

Bernoulli dengan pendekatan poisson. Metode ini termasuk salah satu model dalam *Divergence From Randomness* model

Dengan metode yang akan diimplementasikan dalam program, pengguna tidak perlu mencari dokumen dengan membaca satu per satu. Pengguna hanya memasukkan sebuah inputan berupa kata kunci yang berkaitan dengan dokumen yang dicari. Diharapkan proses *retrieval* dokumen dengan model acak Bernoulli dengan pendekatan Poisson ini dapat menghasilkan perankingan dokumen yang akurat dan relevan sesuai kata kunci yang dimasukkan.

1.2 Rumusan Masalah

Penelitian ini akan mengamati dan mengevaluasi kinerja model DFR PL2 menggunakan *interpolated precision* dalam pencarian dokumen yang dibutuhkan oleh pengguna.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

- a. Program hanya dapat menghitung relevansi dokumen teks yang berekstensi .txt saja.
- b. Tidak dilakukan proses stemming terhadap term.
- c. Kata-kata yang termasuk daftar *stopword* pada *query* dan dokumen akan dibuang karena kata tersebut kurang bermakna, *stopword* menggunakan bahasa Indonesia.
- d. *Query* tidak mengenal operator logika dan operator bantuan lainnya (misal:tanda “ ”,+,-,AND, OR, dan sebagainya).
- e. Adanya pembatasan panjang *query* yang diinputkan yaitu maksimal 7 kata.
- f. Pencarian dokumen berdasarkan keyword yang context-free, sehingga tidak bergantung konteks tertentu dalam bahasa tertentu.

1.4 Tujuan

Tujuan utama penelitian ini adalah mengimplementasikan serta meneliti keakuratan sistem temu kembali informasi dengan metode formula PL2 (model acak Bernoulli dengan pendekatan Poisson) untuk mencari dokumen yang relevan sesuai kata kunci yang diinputkan oleh pengguna.

1.5 Metode/Pendekatan

Ada 3 jenis tahapan metode yang akan digunakan dalam penelitian ini. Pertama adalah metode pengumpulan data. Penulis melakukan studi pustaka untuk mengumpulkan data. Studi pustaka tersebut dilakukan dengan mempelajari teori serta literatur yang mendukung penelitian ini terutama yang berhubungan dengan distribusi binomial dan poisson, *Information retrieval*, metode *probabilistic model* termasuk didalamnya *Divergence From Randomness*, serta perangkat lunak yang digunakan untuk membangun sistem. Penulis juga melakukan pengumpulan data berupa mengumpulkan dokumen .txt yang digunakan untuk penelitian ini, didapat dari salinan artikel berita berbagai topik yang diambil dari website kompas dan membuat daftar *stop word* yang akan dipakai.

Metode selanjutnya yaitu metode pengembangan sistem. Metode yang dipakai untuk mengembangkan sistem yang dipakai adalah model acak Bernoulli dengan pendekatan Poisson atau yang disebut formula PL2. Metode ini merupakan bagian dari *Divergence From Randomness* model.

Metode yang terakhir adalah metode evaluasi. Metode evaluasi dilakukan dengan menghitung nilai *Interpolated Precision* dan *recall* dari proses *retrieval* yang dilakukan. Nilai *precision* dan *recall* inilah yang menentukan berhasil atau tidaknya sistem yang dibangun menggunakan model acak Bernoulli dengan pendekatan Poisson ini.

1.6 Sistematika Penulisan

Adapun sistematika laporan tugas akhir ini dibagi menjadi 5 bagian, yaitu:

Bab 1 Pendahuluan berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan laporan

Bab 2 Tinjauan Pustaka, bab ini berisi dasar teori yang diperlukan dalam penulisan tugas akhir ini seperti teori tentang distribusi binomial dan poisson, *information retrieval*, *probabilistic retrieval* dan lain-lain.

Bab 3 Perancangan Sistem berisi gambaran umum perancangan perangkat lunak yang dikembangkan. Didalamnya 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, bagian ini akan membahas implementasi perancangan perangkat lunak ke dalam kode program. Bab ini juga berisi analisis dari perangkat lunak yang dikembangkan.

Bab 5 Kesimpulan, merupakan bagian akhir dari tugas akhir ini. Berisi kesimpulan dan saran dari pembuatan tugas akhir ini.



BAB 5

Kesimpulan dan Saran

5.1 Kesimpulan

Dalam penelitian ini, pengujian dilakukan terhadap 118 korpus dokumen, maka dapat diambil kesimpulan sebagai berikut:

1. Kinerja model DFR PL2 cukup baik dalam pencarian dokumen. Hal ini ditunjukkan dengan nilai rata-rata *interpolated precision* yang tinggi yaitu 92,24%.
2. Metode ini tidak dapat menemukan dokumen relevan yang tidak mengandung term query.

5.2 Saran

1. Pencarian dokumen bertipe *.doc atau *.pdf. Pencarian dokumen bertipe *.doc atau *.pdf akan menambah informasi yang dibutuhkan pengguna karena pengguna juga banyak menggunakan tipe data tersebut dalam menyimpan suatu informasi.
2. Penggunaan operator boolean dalam pencarian dokumen. Diharapkan dengan adanya operator boolean, hasil dari retrieval sistem lebih akurat. Penggunaan operator boolean akan meminimalisir dokumen tidak relevan yang diambil oleh sistem.
3. Pengembangan sistem agar proses indexing dokumen yang lebih cepat. Proses indexing yang cepat akan membuat penambahan data lebih banyak tanpa memerlukan waktu yang lama.

DAFTAR PUSTAKA

- Boediono, Koster, A. (2001). *Teori dan Aplikasi: Statistika dan Probabilitas*. Bandung: PT Remaja Rosdakarya
- David A. Grossman, Ophir Frieder.(2004). *Information Retrieval: Algorithms And Heuristics*, 2nd edition, Springer.
- G. Amati and C. J. van Rijsbergen. *Probabilistic models of information retrieval based on measuring the divergence from randomness*. In ACM Transactions on Information Systems (TOIS), volume 20(4), pages 357 - 389, 2002.
- He, Ben and Ounis, Iadh. *A Query-based Pre-retrieval Model Selection Approach to Information Retrieval*. United Kingdom: Department of Computing Science University of Glasgow.
- He, Ben and Ounis, Iadh. *Term Frequency Normalisation Tuning for BM25 and DFR Models*. United Kingdom: Department of Computing Science University of Glasgow.
- Mandala, Rila & Setiawan, Hendra. *Peningkatan Performansi Sistem Temu-Kembali Informasi dengan Perluasan Query Secara Otomatis*. Bandung: Laboratorium Keahlian Informatika Teori Departemen Teknik Informatika ITB
- Manning, D.Christopher, Raghavan, Prabhakar & Schütze, Hinrich. (2008). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press
- Neto, Berthier Ribeiro & Yates, Ricardo Baeza. (2006). *Modern Information Retrieval*. Reading, Massachusetts: Addison Wesley
- Priyanto Rahmat. (2009). *Langsung Bisa Visual Basic .NET 2008*, Andi Offset, Yogyakarta