

**IMPLEMENTASI METODE LANGUAGE MODEL DALAM  
SISTEM PENCARIAN ARTIKEL BERBAHASA KOREA**

Skripsi



oleh  
**RIEKA INDAH PERMATASARI**  
**22094765**

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

# **IMPLEMENTASI METODE LANGUAGE MODEL DALAM SISTEM PENCARIAN ARTIKEL BERBAHASA KOREA**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana  
Sebagai Salah Satu Syarat dalam Memperoleh Gelar  
Sarjana Komputer

Disusun oleh

**RIEKA INDAH PERMATASARI**  
**22094765**

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

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### IMPLEMENTASI METODE LANGUAGE MODEL DALAM SISTEM PENCARIAN ARTIKEL BERBAHASA KOREA

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 Mei 2013



RIEKA INDAH PERMATASARI  
22094765

## HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI METODE LANGUAGE MODEL  
DALAM SISTEM PENCARIAN ARTIKEL  
BERBAHASA KOREA

Nama Mahasiswa : RIEKA INDAH PERMATASARI

N I M : 22094765

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

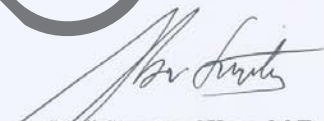
Semester : Genap

Tahun Akademik : 2012/2013

© UKDW

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 28 Mei 2013

Dosen Pembimbing I

  
Budi Susanto, SKom.,M.T.

Dosen Pembimbing II

  
Dra. Widi Hapsari, M.T.

## HALAMAN PENGESAHAN

### IMPLEMENTASI METODE LANGUAGE MODEL DALAM SISTEM PENCARIAN ARTIKEL BERBAHASA KOREA

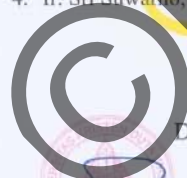
Oleh: RIEKA INDAH PERMATASARI / 22094765

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 22 Mei 2013

Yogyakarta, 28 Mei 2013  
Mengesahkan,

Dewan Penguji:

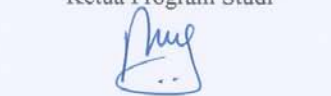
1. Budi Susanto, S.Kom., M.T.
2. Dra. Widi Hapsari, M.F.
3. Erick Purwanto, S.Kom, M.Com.
4. Ir. Sri Suwarno, M.Eng.



Dekan

  
(Drs. Wimmie Handiwidjojo, M.T.)

Ketua Program Studi

  
(Nugroho Agus Haryono, M.Si)

## UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, yang telah memberikan rahmat dan anugerah serta penyertaan yang sempurna, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul Implementasi Metode *Language Model Query Likelihood* dalam Sistem Pencarian Artikel berbahasa Korea.

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

Dalam proses penyelesaian dan pembuatan program serta laporan Tugas Akhir ini, penulis telah banyak menerima bimbingan, masukan, saran dan dukungan dari berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada :

1. Bapak Budi Susanto, S.Kom. M.T. selaku dosen pembimbing I yang telah banyak memberikan bimbingan, petunjuk, saran, masukan dan dukungan dengan sabar sekali dan hangat kepada penulis dari awal pembuatan Tugas Akhir sampai akhir.
2. Bapak Dra. Widi Hapsari, M.T. selaku dosen pembimbing II yang telah banyak memberikan bimbingan, petunjuk, saran, masukan dan dukungan dengan sabar dan hangat kepada penulis dari awal pembuatan Tugas Akhir sampai akhir.
3. Dosen-dosen Universitas Kristen Duta Wacana yang juga telah membantu memberikan bimbingan, petunjuk, saran, masukan dan dukungan kepada penulis.
4. Papa, Mama, kakak, dan adik tercinta yang tak henti-hentinya memberikan doa, dukungan, semangat, motivasi dan dorongan baik secara materiil maupun moril sehingga Tugas Akhir ini dapat terselesaikan dengan baik dan tepat waktu.

5. Para sahabat seperjuangan dalam mengerjakan skripsi: Fanny, Friska, Putri, Ria, Ela, Anton, Patria, Evline, Kak Michell, Lina dan teman seperjuangan lainnya yang sama-sama mengambil Tugas Akhir terkhusus angkatan 2009, yang berjuang bersama untuk menyelesaikan Tugas Akhir ini dengan saling memotivasi dan mengingatkan serta saling memberikan dukungan.
6. Para sahabat tercinta dan tersayang : Fanny, Friska, Rosi, Putri, Ria, Diana, Thirza, Selvi dan para sahabat lainnya yang selalu setia dan tulus memberikan doa, dukungan, semangat, motivasi, dan dorongan sehingga Tugas Akhir ini dapat terselesaikan dengan baik dan tepat waktu.
7. Pihak lain yang tidak dapat penulis sebutkan satu per satu, sehingga Tugas Akhir ini dapat terselesaikan dengan baik dan tepat waktu.

Penulis menyadari bahwa program dan laporan dalam Tugas Akhir ini masih jauh dari pada kata sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang dapat membangun dari para pembaca sekalian, sehingga dalam waktu kedepan, penulis dapat memberikan karya yang lebih baik lagi.

Akhir kata, penulis ingin meminta maaf yang sebesar-besarnya apabila dalam proses pembuatan program dan penyusunan laporan Tugas Akhir ini terselip kata-kata yang salah maupun kesalahan yang pernah penulis lakukan. Semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Yogyakarta, 14 Mei 2013

Penulis

## INTISARI

### Implementasi Metode Language Model Query Likelihood dalam Sistem Pencarian Artikel menggunakan Bahasa Korea

Seiring dengan berkembangnya teknologi saat ini, semakin banyak informasi yang tersedia dalam bentuk dokumen teks. Dokumen teks dapat berisi kumpulan teks yang menggunakan berbagai macam *character set*. *Character set* ini dapat digunakan sebagai representasi dari bahasa yang digunakan, sebagai contoh *character set* unicode untuk bahasa Korea. Jika sistem temu kembali (IR) yang berkembang saat ini mampu memproses & menyediakan fungsi pencarian terhadap kumpulan dokumen, seharusnya juga dapat diterapkan untuk kumpulan dokumen yang mengandung berbagai *character set*.

Melihat latar belakang di atas, penulis membangun sebuah sistem yang dapat melakukan pencarian artikel dalam bahasa Korea, dimana *user* menginputkan kata kunci yang akan dicari menggunakan karakter hangul sehingga dapat menghasilkan daftar artikel yang sesuai dengan kata kunci dengan menggunakan metode *Language Model Query Likelihood*.

Sistem pencarian artikel dengan metode *Language Model Query Likelihood* pada sistem yang telah dibangun ini, memberikan nilai rata – rata *precision* yang masih cukup tinggi untuk hasil uji coba yang telah dilakukan, yakni sebesar 0.72 dari nilai rata – rata *precision* per query. Metode *Language Model Query Likelihood* yang digunakan tidak ada kaitannya dengan karakter unicode korea terkait dengan *information retrieval* karena metode ini hanya menghitung probabilitas setiap token pada setiap dokumen dimana nilai probabilitas tersebut tidak dipengaruhi apakah token tersebut berupa karakter unicode atau tidak.

Kata kunci : korea, *Language Model Query Likelihood*, pencarian, *precision*, *recall*, unicode



## DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	viii
DAFTAR ISI.....	ix
DAFTAR TABLE.....	xii
DAFTAR GAMBAR DAN GRAFIK.....	xiv
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 Metode Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB 2 LANDASAN TEORI.....	5
2.1 Perumusan Masalah.....	5
2.2 Landasan Teori.....	6
2.2.1 Karakter Unicode Korea.....	6
2.2.2 <i>Information Retrieval</i> .....	6
2.2.3 <i>Indexing</i> .....	6
2.2.4 Language Model.....	7
2.2.1 <i>Precision</i> dan <i>Recall</i> .....	16
BAB 3 ANALISIS DAN PERANCANGAN SISTEM.....	20
3.1 Kebutuhan Fungsional dan Non Fungsional.....	20
3.2 Algoritma dan Flowchart.....	22
3.2.1 <i>Indexing</i> dengan <i>Inverted Index</i> .....	22

3.2.2 Pencarian dengan Metode Query-Likelihood.....	25
3.3 Kamus Data .....	26
3.3.1 Tabel dokumen_training.....	26
3.3.2 Tabel token_training.....	26
3.3.3 Tabel dokumen_token_training.....	27
3.3.4 Tabel stopword .....	27
3.3.5 Tabel perhitungan .....	27
3.3.6 Tabel token_testing .....	29
3.3.7 Tabel user .....	29
3.4 Diagram Skema .....	30
3.5 Rancangan Antarmuka Sistem.....	30
3.6 Rancangan Pengujian Sistem.....	32
<b>BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....</b>	<b>39</b>
4.1 Implementasi Sistem.....	39
4.1.1 Blok Diagram Sistem .....	39
4.1.2 Antarmuka Sistem .....	40
A. Halaman Home untuk Admin .....	40
B. Halaman Daftar Dokumen <i>Training</i> .....	41
C. Halaman Edit Dokumen .....	42
D. Halaman Tambah Dokumen <i>Training</i> .....	43
E. Halaman Tambah Stopword .....	44
4.1.3 Pseudocode Sistem .....	44
A. Pelatihan .....	44
B. Pencarian .....	52
4.2 Analisis Sistem .....	53
4.2.1 Pengujian Pertama .....	54
4.2.2 Pengujian Kedua.....	54
4.2.3 Pengujian Ketiga .....	56
4.2.4 Pengujian Keempat.....	57
4.2.5 Pengujian Kelima .....	58
4.2.6 Pengujian Keenam.....	59

4.2.7 Pengujian Ketujuh .....	60
4.2.8 Pengujian Kedelapan .....	61
4.2.9 Pengujian Kesembilan .....	62
4.2.10 Pengujian Kesepuluh .....	63
4.2.11 Rata – rata Hasil Pengujian .....	65
4.2.12 Grafik Interpolasi.....	66
BAB 5 KESIMPULAN DAN SARAN .....	67
5.1 Kesimpulan.....	67
5.2 Saran .....	67
DAFTAR PUSTAKA .....	68

©UKDW

## DAFTAR TABEL

Tabel 2.1 Tabel yang berisi jumlah dokumen yang berisi tiap <i>token</i> .....	11
Tabel 2.2 Tabel yang berisi jumlah setiap <i>token</i> di seluruh dokumen .....	11
Tabel 2.3 Tabel yang berisi jumlah setiap <i>token</i> dalam setiap dokumen .....	11
Tabel 2.4 Tabel <i>maximum likelihood</i> untuk setiap <i>token</i> .....	12
Tabel 2.5 Tabel rata – rata probabilitas untuk setiap <i>token</i> .....	13
Tabel 2.6 Tabel nilai <i>mean term frequency</i> tiap token dalam tiap dokumen .....	13
Tabel 2.7 Tabel nilai <i>risk value</i> untuk setiap <i>token</i> dalam setiap dokumen .....	14
Tabel 2.8 Tabel probabilitas setiap <i>token</i> dalam setiap dokumen .....	15
Tabel 2.9 Tabel Similarity dengan Query Likelihood .....	16
Tabel 2.10 Tabel ketergantungan untuk menghitung nilai precision dan recall ...	17
Tabel 3.1 Tabel dokumen_training .....	26
Tabel 3.2 Tabel token_training .....	26
Tabel 3.3 Tabel dokumen_token_training .....	27
Tabel 3.4 Tabel stopwords .....	27
Tabel 3.5 Tabel perhitungan .....	28
Tabel 3.6 Tabel token_testing .....	29
Tabel 3.7 Tabel user .....	29
Tabel 3.8 Tabel dum_index .....	30
Tabel 3.9 Tabel Query dan Dokumen yang Relevan .....	33
Tabel 4.1 Tabel Hasil Pengujian Pertama .....	45
Tabel 4.2 Tabel Hasil Pengujian Kedua .....	46
Tabel 4.3 Tabel Hasil Pengujian Ketiga .....	47
Tabel 4.4 Tabel Hasil Pengujian Keempat .....	48
Tabel 4.5 Tabel Hasil Pengujian Kelima .....	49
Tabel 4.6 Tabel Hasil Pengujian Keenam .....	50
Tabel 4.7 Tabel Hasil Pengujian Ketujuh .....	51
Tabel 4.8 Tabel Hasil Pengujian Kedelapan .....	52

Tabel 4.9 Tabel Hasil Pengujian Kesembilan .....	53
Tabel 4.10 Tabel Hasil Pengujian Kesepuluh .....	54
Tabel 4.11 Tabel Rata – rata Hasil Pengujian .....	55

©UKDW

## DAFTAR GAMBAR & GRAFIK

Gambar 2.1 Contoh Karakter Unicode Korea .....	6
Gambar 2.2 Arsitektur Query Likelihood Model.....	8
Gambar 2.3 Kurva <i>precision / recall</i> untuk <i>ranked retrieval</i> .....	18
Gambar 2.4 Kurva <i>averaged 11 poin precision / recall</i> .....	19
Gambar 3.1 Diagram <i>Use Case</i> .....	21
Gambar 3.2 <i>Flowchart preprocessing</i> .....	23
Gambar 3.3 <i>Flowchart indexing</i> .....	24
Gambar 3.4 <i>Flowchart</i> Perhitungan Query Likelihood .....	25
Gambar 3.5 Diagram Skema .....	30
Gambar 3.6 Halaman Utama untuk Sistem Pencarian Artikel.....	30
Gambar 3.7 Halaman Utama untuk Admin .....	31
Gambar 3.8 Halaman Upload Dokumen Training .....	31
Gambar 4.1 Blok Diagram Indexing .....	36
Gambar 4.2 Blok Diagram Retrieval .....	37
Gambar 4.3 Implementasi Halaman Pencarian .....	37
Gambar 4.4 Implementasi Halaman Daftar Dokumen Training .....	38
Gambar 4.5 Implementasi Halaman Edit Dokumen .....	39
Gambar 4.6 Implementasi Halaman Dokumen Training .....	40
Gambar 4.7 Implementasi Halaman Penambahan Stopword .....	41
Grafik 4.1 Grafik rata – rata <i>precision</i> hasil pengujian .....	63

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Seiring dengan berkembangnya teknologi saat ini, semakin banyak informasi yang tersedia dalam bentuk dokumen teks. Dokumen teks dapat berisi kumpulan teks yang menggunakan berbagai macam *character set*. *Character set* ini dapat digunakan sebagai representasi dari bahasa yang digunakan, sebagai contoh *character set* unicode untuk bahasa Korea. Jika sistem temu kembali (IR) yang berkembang saat ini mampu memproses & menyediakan fungsi pencarian terhadap kumpulan dokumen, seharusnya juga dapat diterapkan untuk kumpulan dokumen yang mengandung berbagai *character set*.

Sistem temu kembali yang berkembang saat ini sebenarnya tidak mempedulikan tentang *character set* yang digunakan dalam kumpulan dokumen. Namun demikian, ada beberapa metode IR yang dapat memberikan fungsi yang memperhatikan tentang model bahasa yang terkandung dalam kumpulan dokumen. Metode IR tersebut adalah *Vector Space Model*, *Simple Term Weight*, *Extended Boolean*, *Query Likelihood Language Model*, dan lain-lain.

Dalam hal ini metode *Query Likelihood* dipandang mampu untuk membantu dalam pembuatan sistem pencarian. *Language Model* ini dalam *information retrieval* difokuskan pada efektifitas secara empiris melalui pemakaian model yang simpel. Salah satu model pendekatan yang dapat digunakan adalah *query-likelihood* yang diperkenalkan pertama kali oleh Ponte dan Croft (1998). Sistem ini akan memberikan nilai relevansi yang tinggi pada dokumen yang memiliki kemiripan paling besar terhadap *query* yang diinputkan oleh *user*. Nilai relevansi yang dihasilkan dalam perhitungan menggunakan *query-likelihood* ini hampir sama efektifnya dengan menggunakan perhitungan *tf-idf*.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang masalah diatas, penulis akan merancang dan membangun sebuah sistem yang akan melakukan proses pencarian dokumen berbahasa Korea tersebut, dengan permasalahan yang akan diuji sebagai berikut :

- Sejauh mana kinerja yang diberikan ke sistem information retrieval menggunakan *query likelihood* untuk dokumen dengan karakter set korea ditinjau dari evaluasi *precision* dan *recall*
- Apakah masih mungkin *language model* dikaitkan dengan unicode untuk karakter korea terkait dengan information retrieval

## 1.3 Batasan Masalah

Untuk menghindari meluasnya pokok bahasan, maka penulis membatasi perumusan masalah sebagai berikut :

- Data yang digunakan adalah dokumen teks berbahasa Korea yang menggunakan karakter Hangul.
- Dokumen teks yang akan digunakan dikumpulkan secara manual dari situs berita <http://healthcare.joinsmsn.com>
- Bentuk masukan ke sistem berupa karakter unicode hangul
- Tidak adanya tahap *stemming* dalam *preprocessing* dokumen
- Kata – kata yang termasuk dalam *stopword-list* pada *query* akan dibuang karena kata – kata tersebut terlalu umum
- Panjang *query* dibatasi hanya 3 – 5 kata

## 1.4 Tujuan Penelitian

Melalui penelitian ini, maka tujuan yang ingin dicapai oleh penulis adalah sebagai berikut :

- Untuk mengetahui seberapa baik sebuah sistem IR terhadap kumpulan dokumen yang mengandung karakter set bahasa korea dengan menggunakan *query likelihood language model*.



- Membantu *user* untuk mendapatkan artikel yang sesuai dengan kata kunci yang diinputkan oleh *user*, dalam hal ini khususnya untuk orang-orang yang tertarik dengan berita dari negara Korea.

## 1.5 Metode Penelitian

Metode yang digunakan oleh penulis dalam penelitian Tugas Akhir ini antara lain :

- Studi pustaka dan literatur  
Studi pustaka dilakukan dengan mencari dan mempelajari sumber – sumber pustaka yang berkaitan dengan pencarian informasi serta mempelajari struktur bahasa korea. Sumber – sumber tersebut berupa buku serta sumber online di internet yang dapat dipercaya.
- Pengembangan Sistem  
Tahap ini dilakukan perancangan basis data untuk menyimpan data bahan penelitian, perancangan antarmuka, dan perancangan prosedural sistem.
- Pengujian  
Pengujian sistem dilakukan untuk mendapatkan hasil dan kesimpulan dari penelitian ini.

## 1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini dikelompokkan menjadi beberapa bab, yaitu :

Bab I, Pendahuluan. Bab ini berfungsi untuk memberikan gambaran umum mengenai penelitian yang akan dilakukan yang mencakup latar belakang, perumusan masalah, batasan masalah, tujuan penelittian, metode penelitian, dan sistematika penulisan penelitian.

Bab II, Landasan Teori. Bab ini berisi mengenai tinjauan pustaka dan landasan teori yang dipakai dalam penelitian ini. Tinjauan pustaka menguraikan berbagai teori yang didapat dari berbagai sumber terkait dengan penelitian ini.

Bab III, Perancangan Sistem. Bab ini berisi mengenai rancangan sistem yang akan dibangun dalam penelitian ini. Rancangan sistem yang akan dibuat berupa spesifikasi dari sistem, rancangan diagram sistem, rancangan antar muka sistem berupa input dan output, serta tahapan yang akan dilakukan dalam perancangan dan pembuatan sistem.

Bab IV, Implementasi dan Analisis Sistem. Bab ini membahas tentang implementasi sistem yang telah dirancang pada bab III dan berisi hasil capture dari sistem yang telah dibuat. Juga terdapat analisa mengenai yang digunakan dalam pembuatan sistem.

Bab V, Kesimpulan dan Saran. Bab ini berisi tentang kesimpulan dari hasil penelitian yang telah dilakukan dan saran – saran yang mungkin dapat dilakukan di masa yang akan datang untuk mengembangkan sistem ini ke arah yang lebih baik, serta dapat menjadi referensi bagi pembaca yang lainnya.

©UKIDN

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan hasil analisis dan implementasi sistem, maka diperoleh kesimpulan sebagai berikut :

1. Dari keseluruhan rata – rata hasil pengujian didapat nilai *precision* 0.72, yang artinya sistem masih cukup baik dalam mengembalikan dokumen – dokumen yang relevan.
2. Tidak adanya kaitan antara metode *language model* dengan karakter unicode korea terkait dengan *information retrieval* karena metode ini hanya memperhatikan probabilitas dari setiap token pada setiap dokumen dan probabilitasnya tidak dipengaruhi token tersebut berupa karakter unicode atau tidak.

#### 5.2. Saran

Saran untuk pengembangan dan perbaikan sistem ini adalah :

1. Perlu adanya perbaikan untuk mempercepat proses pencarian dengan meng-optimalisasi proses *indexing*-nya.
2. Penggunaan kamus bahasa Korea untuk melakukan pengecekan dokumen bahasa Korea serta penambahan proses *stemming*.

## DAFTAR PUSTAKA

- Aloufi, Khalid S. R. (2011). Diacritic Oriented Arabic Information Retrieval System. Diakses tanggal 6 Februari 2013 dari <http://cscjournals.org/csc/manuscript/Journals/IJCSS/volume5/Issue1>
- Grossman, D. A., & Frieder, O. (2007). *Information Retrieval Algorithms and Heuristics Second Edition*. Chicago, IL, U.S.A: Illinois Institute of Technology.
- Hiemstra, Djoerd., Robertson, S., Zaragoza, H. Parsimonious Language Models for Information Retrieval. Diakses tanggal 4 Februari 2013 dari [http://research.microsoft.com/pubs/66933/hiemstra\\_sigir04.pdf](http://research.microsoft.com/pubs/66933/hiemstra_sigir04.pdf).
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Swapna, N., Kumar, N. H., Rani, B. P. (2012). INFORMATION RETRIEVAL IN INDIAN LANGUAGES: A CASE STUDY ON CROSS LINGUAL AND MULTI LINGUAL. Diakses tanggal 8 Februari 2013 dari <http://www.ijrcct.org/index.php/ojs/article/view/52>

©UKYD