

**Model Indeks Basis Data untuk Peningkatan Performa
Mesin Pencari menggunakan Bahasa Indonesia**

TESIS

Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer (MKOM)



Oleh:

Agung Riyadi

1111600167

**POGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

JAKARTA

2013



**PROGRAM STUDI: MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

LEMBAR PENGESAHAN

Nama Mahasiswa : Agung Riyadi
Nomor Induk Mahasiswa : 1111600167
Konsentrasi : Rekayasa Komputer Terapan
Jenjang Studi : Magister Ilmu Komputer
Judul : Model Indeksi Basis Data untuk Peningkatan Performa Mesin Pencari Menggunakan Bahasa Indonesia

Jakarta, 09 Pebruari 2013

Tim Penguji:

Ketua,

Dr. Ir. Nazori Az, M.T

Anggota,

Dr. Moedjiono, M.Sc

Pembimbing Utama,

Dr. Dwi handoko

Tanda Tangan:

[Handwritten signatures of the examiners]

Ketua Program Studi

[Handwritten signature of Dr. Ir. Nazori Az, M.T]
Dr. Ir. Nazori Az, M.T

LEMBAR PERNYATAAN



PROGRAM STUDI MAGISTER KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Agung Riyadi
NIM : 1111 600167
Program Studi : Magister komputer
Program : Pekayasa komputer Terapan

menyatakan bahwa TESIS yang berjudul:

Model Indeks Basis Data untuk peningkatan
Performa Mesin Pencari menggunakan Bahasa
Indonesia

1. Merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain,
2. Saya ijin untuk dikelola oleh Universitas Budi Luhur sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila di kemudian hari pernyataan ini tidak benar.

Jakarta, 9 Pebruari 2013



(Agung Riyadi)

ABSTRAK

Mesin pencari yang sudah ada dalam menampilkan hasil pencarian hanya menggunakan kepopuleran sebuah situs saja. Oleh karena itu, akurasi dalam penyajian hasil pencarian sangat rendah. Dengan menggunakan sistem indeks, sebuah mesin pencari dapat melakukan proses pencarian lebih akurat dan cepat dalam menyajikan hasil pencarian. Pada beberapa sistem indeks dari penelitian sebelumnya, nilai akurasi dari hasil pencarian masih kurang terutama untuk artikel khusus berbahasa indonesia. Dalam penelitian ini dilakukan sistem pengindeksan artikel khusus bahasa indonesia, dengan teknik ekstraksi kata kunci menggunakan metode hybrid menggunakan metode dictionary matching dan TFIDF. Penelitian ini bertujuan untuk membuat sistem indeks database untuk artikel khusus bahasa indonesia, yang nantinya akan digunakan untuk proses pencarian pada mesin pencari. Berdasarkan hasil pengujian, nilai akurasi ekstraksi kata kunci dengan review manusia sebesar 19,83%. Nilai akurasi penyajian hasil pencarian sebesar 73,92% berdasarkan review manusia. Hasil ini lebih besar jika dibandingkan dengan mesin indeks DBMS, dan indeks akselerator. Untuk kecepatan rata-rata hasil pencarian dengan sistem indeks ini adalah 56,40 artikel permili-sekon.

Keywords: Teks prosesing, Ekstraksi kata kunci, indeks database, TFIDF, Search Engine

ABSTRACT

The existing search engine only show the search results based on the web popularity. Because of that, the accuracy in search result presentation very low. Using indexing database systems, a search engine process can be more accurate and faster. In some indexing method of previous research, the accuracy value from search engine still low, especially for indonesian language article. This research describe about indonesian article indexing using keyword extraction based on hybrid method using dictionary matching and TFIDF. The purpose of this research to develop a database indexing system contains of indonesian article especially. Furthermore this model will be supporting several existing search engine applications. Based on the testing results, the accuracy of keyword extraction by human review is 19.83%. The accuracy value of search results by human review is 73.93%. This result is higher than other indexing systems like DBMS indexing, and using accelerator indexing. The average speed for the presentation of search result is 56.40 articles per milisecond.

Keywords: *Text Processing, Keywords Extraction, Database Indexing, TFIDF, Search Engine*

KATA PENGANTAR

Alhamdulillah, segala puji dan syukur kepada Allah SWT yang telah memberikan kekuatan dan kesehatan sehingga penulis dapat menyelesaikan penyusunan tesis ini. Shalawat dan salam selalu tercurahkan kepada junjungan kita Nabi Muhammad SAW.

Dalam kesempatan ini, penulis juga ingin mengucapkan rasa terima kasih kepada:

1. Ibu dan ayahku yang selalu mendoakan anaknya di setiap waktu demi kehidupan yang terbaik untuk anaknya.
2. Bapak Dr. Dwi Handoko selaku dosen pembimbing sehingga tesis ini dapat terselesaikan dengan baik.
3. Bapak Dr. S. Moedjiono, M.Sc, selaku direktur program pascasarjana Universitas Budiluhur.
4. Bapak Dr. Ir. Nazori AZ, MT, selaku ketua program studi Magister Komputer Universitas Budiluhur.
5. Teman-teman mahasiswa pascasarjana magister komputer X-A angkatan 2011, yang telah banyak membantu selama perkuliahan.
6. Kepada Zissalwa Hafsari, yang selalu men-support dan mendoakanku dalam perkuliahan hingga penulisan tesis ini terselesaikan.
7. Kepada para reviewer, Yuni arti, Ninon N. Faiza, Intan Permatasari, Dian kurniawan, dan lainnya. Yang ikut membantu proses pengujian dalam penelitian ini.

Dengan rasa tulus dalam hati, penulis menyampaikan banyak terima kasih dan permohonan maaf pada pihak yang tidak dapat disebutkan satu persatu. Semoga Allah SWT. Melimpahkan balasan yang sesuai dengan segala yang telah diberikan dan semoga tesis ini dapat sangat bermanfaat bagi para pembacanya. Amin.

DAFTAR ISI

	Halaman
Lembar Pengesahan	i
ABSTRAK	ii
ABSTRACT	iii
KATA PENGANTAR	iv
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
DAFTAR LAMPIRAN	ix
 BAB I PENDAHULUAN	 1
1.1 Latar belakang	1
1.2 Masalah Penelitian.....	3
1.2.1 Identifikasi Masalah	3
1.2.2 Batasan Masalah.....	3
1.2.3 Rumusan masalah.....	3
1.3 Tujuan dan Manfaat Penelitian	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	4
1.4 Ruang Lingkup	4
1.5 Sistematika/Tata Urut Penulisan.....	4
1.6 Daftar Pengertian/Istilah.....	5
 BAB II LANDASAN TEORI DAN KERANGKA KONSEP	 8
2.1 Tinjauan Pustaka	8
2.1.1 Stopwords	8
2.1.2 Stemming	9
2.1.3 Ekstraksi kata kunci	10
2.1.4 Metode indeks kata kunci	11
2.1.4.1 Pendekatan statistik sederhana	11
2.1.4.2 Pendekatan lingustik.....	12
2.1.4.3 Pendekatan <i>machine learning</i>	12
2.1.4.4 Pendekatan lainnya	12
2.2 Tinjauan Studi	12
2.2.1 Penelitian yang dilakukan oleh Xiaochun Yang et all... ..	12
2.2.2 Penelitian yang dilakukan oleh Feifan Liu et all	13
2.2.3 Penelitian yang dilakukan oleh Qi Shu dan Jennifer	13
2.2.4 Perbandingan beberapa penelitian yang terkait	14
2.3 Tinjauan objek penelitian	16
2.4 Kerangka Pemikiran.....	16
2.5 Hipotesis.....	17

BAB III DESAIN PENELITIAN.....	18
3.1 Metodologi Penelitian	18
3.2 Metode yang diajukan	19
3.3 Pemilihan Sampel.....	20
3.4 Metode Pengumpulan data	21
3.5 Pengujian Hipotesis.....	22
3.6 Instrumentasi	23
3.7 Teknik Analisis Data	24
3.7.1 Perhitungan perbandingan kata kunci	24
3.7.2 Perhitungan perbandingan kecepatan proses	24
3.8 Langkah-langkah Penelitian	25
3.9 Jadwal Penelitian	26
BAB IV PEMBAHASAN HASIL PENELITIAN.....	27
4.1 Pengelompokan dan Analisis Data.....	27
4.2 Temuan dan Interpretasi.....	28
4.3 Rancangan Sistem dan Prototype Model	30
4.3.1 Rancangan database artikel dan database indeks	31
4.3.2 Rancangan mesin crawler.....	33
4.3.3 Rancangan aplikasi ekstraksi dan indeks kata kunci.....	34
4.4 Pengujian Sistem.....	36
4.4.1 Pengujian Akurasi	36
4.4.2 Pengujian Akurasi Hasil Pencarian.....	38
4.4.3 Pengujian Kecepatan Hasil Pencarian	40
4.5 Implikasi penelitian.....	42
4.5.1 Aspek Sistem	43
4.5.1.1 Hardware	43
4.5.1.2 Software	43
4.5.1.3 Arsitektur	44
4.5.2 Aspek Penelitian Lebih lanjut.....	44
4.6 Rencana Implementasi	45
BAB V PENUTUP.....	46
4.1 Kesimpulan	46
4.2 Saran	46
DAFTAR PUSTAKA.....	47
LAMPIRAN-LAMPIRAN	55
RIWAYAT HIDUP.....	75

DAFTAR GAMBAR

Gambar	Halaman
II.1 Contoh penggunaan proses stopwords	9
II.2 Contoh penggunaan proses stemming	10
II.3 Arsitektur Sistem Indeksi [Qi and Jennifer 2005]	14
II.4 Kerangka Pemikiran penelitian	17
III.1 Metodologi Penelitian	18
III.2 Proses Indeksi	20
III.3 Flowchart metode pengumpulan data artikel dari internet	21
III.4 Proses pengumpulan data	21
III.5 Metode Pengujian Akurasi Perbandingan Automatic Keyword Extraction dengan Human Keyword extraction	22
III.6 Metode Pengujian Kecepatan Pencarian dengan Indeksi DBMS, Indeks akselerator, dan indeksi dengan metode hybrid.	23
III.7 Model Vector Space untuk Review Keyword	24
III.8 Langkah Penelitian sistem pengindeksian.	25
IV.1 Rancangan sistem indeksi	30
IV.2 Relasi tabel artikel dan tabel indeksi.....	32
IV.3 Tampilan grafis aplikasi crawler	34
IV.4 Diagram alir ekstraksi kata kunci.....	35
IV.5 Diagram alir indeksi kata kunci	36
IV.6 Grafik perbandingan ekstraksi kata kunci dengan review manusia	38
IV.7 Arsitektur untuk pengembangan sistem indeksi.....	44
IV.8 Aplikasi mesin pencari dengan menggunakan sistem Indeksi.	45

DAFTAR TABEL

Tabel	Halaman
II.1 Ekstraksi kata kunci dan hasil evaluasi pada penelitian sebelumnya	15
III.1 Timeline Kegiatan Penelitian	26
IV.1 Pengukuran modul yang digunakan untuk sistem indeksasi.....	28
IV.2 Hasil pengujian akurasi Ekstraksi kata kunci	37
IV.3 Hasil pencarian dengan menggunakan ketiga indeksasi.....	39
IV.4 Hasil pencarian dengan menggunakan indeks DBMS	40
IV.5 Hasil pencarian dengan menggunakan indeks Akselerator	41
IV.6 Hasil pencarian dengan menggunakan indeks metode hybrid	41
IV.7 Timeline rencana implementasi pembuatan mesin pencarian	45

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Source Code Aplikasi	49
2. Screenshot Aplikasi.....	55
3. Hasil Pengujian.....	59



DAFTAR PUSTAKA

- [Anette, 2003] Anette Hulth. "Enhancing Linguistically Oriented Automatic Keyword Extraction". Dept. of Computer and Systems Sciences Stockholm University SE-164 40 Kista, Sweden.2003.
- [Chengzhi ea 2008] Chengzhi ZHANG, Huilin WANG, Yao LIU, Dan WU, Yi LIAO, Bo WANG . "Automatic Keyword Extraction from Documents Using Conditional Random Fields ". Journal of Computational Information Systems4:3(2008) 1169-1180 . 2008.
- [Iryna, 2009] Iryna Oelze . "Automatic Keyword Extraction for Database Search". Master of Thesis, L3S Research Center University of Hannover, Germany.2009.
- [Jianwei ea 2010] Jianwei Zhang , Yukiko Kawai and Tadahiko Kumamoto . "A Flexible Re-ranking System Based on Sub-keyword Extraction and Importance Adjustment ". IAENG International Journal of Computer Science, 37:3, IJCS_37_3_10.2010.
- [Kian 2011] H. H. Kian, and M. Zahedi. "An Efficient Approach For Keyword Selection; Improving Accessibility Of Web Contents By General Search Engines ". International Journal of Web & Semantic Technology (IJWest) Vol.2, No.4, October 2011.
- [Kuppusamy 2012] Kuppusamy K.S, dan Aghila G. "A Model for Personalized Keyword Extraction from Web Pages using Segmentation ". International Journal of Computer Applications (0975 – 8887) Volume 42– No.4, March 2012 .
- [Liu ea 2009] Feifan Liu, Deana Pennell, Fei Liu and Yang Liu . "Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts ". Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL, pages 620–628.2009.
- [Matsuo 2003] Yutaka Matsuo , dan Mitsuru Ishizuka . "Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information ". National Institute of Advanced Industrial Science and Technology . 2003.