



**ANALISA TINGKAT KECOCOKAN PENCARIAN NAMA  
DALAM BASIS DATA DENGAN MENGGUNAKAN  
ALGORITMA SOUNDEX DAN ALGORITMA JARO-WINKLER**

**IBNU FARID ISKANDAR  
41506110165**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2010**



**ANALISA TINGKAT KECOCOKAN PENCARIAN NAMA  
DALAM BASIS DATA DENGAN MENGGUNAKAN  
ALGORITMA SOUNDEX DAN ALGORITMA JARO-WINKLER**

*Laporan Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

IBNU FARID ISKANDAR  
41506110165

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2010

## LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

NIM : 41506110165  
Nama : IBNU FARID ISKANDAR  
Judul Skripsi : ANALISA TINGKAT KECOCOKAN PENCARIAN NAMA  
DALAM BASIS DATA DENGAN MENGGUNAKAN  
ALGORITMA SOUNDEX DAN ALGORITMA JARO-  
WINKLER

Menyatakan bahwa skripsi tersebut di atas adalah hasil karya saya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan skripsi saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 1 Agustus 2010

Ibnu Farid Iskandar

## LEMBAR PERSETUJUAN

NIM : 41506110165  
Nama : IBNU FARID ISKANDAR  
Judul Skripsi : ANALISA TINGKAT KECOCOKAN PENCARIAN NAMA  
DALAM BASIS DATA DENGAN MENGGUNAKAN  
ALGORITMA SOUNDEX DAN ALGORITMA JARO-  
WINKLER

SKRIPSI INI TELAH DIPERIKSA DAN DISETUJUI

JAKARTA, 1 AGUSTUS 2010

Devi Fitriyah, S.Kom., MTI  
Pembimbing

Devi Fitriyah, S.Kom., MTI  
Koord. Tugas Akhir Teknik Informatika

Abdusy Syarif, ST., MT  
KaProdi Teknik Informatika

## HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan khusus kepada umik, ibunda tersayang yang selalu memberikan kasih sayangnya serta telah begitu tegar mengasuh ketiga buah hatinya sebagai *single-parent* sejak meninggalnya ayah saat penulis masih bayi.

## KATA PENGANTAR

Puji syukur Alhamdulillah kehadiran Allah SWT yang telah melimpahkan segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Jurusan Teknik Informatika Universitas Mercu Buana.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa laporan tugas akhir ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Devi Fitriyah, S.Kom., MTI selaku dosen pembimbing tugas akhir dan koordinator tugas akhir pada Jurusan Teknik Informatika Universitas Mercu Buana.
2. Umik tersayang yang telah mendukung penulis baik moril maupun materiil.
3. Isteriku Wardah yang selalu memberikan dukungan penuh dalam menyelesaikan tugas akhir ini. *I love you.*
4. Wafi dan Vina, buah hati kami yang selalu menghadirkan keceriaan pada setiap kebersamaan dan mengundang kerinduan pada saat sebaliknya.
5. Sahabat dan rekan kerja yang telah memberikan dukungan moral untuk menyelesaikan tugas akhir ini.

6. Rekan-rekan seperjuangan angkatan IX atas kebersamaan dan dukungan moral sejak selama masa kuliah hingga dalam penyelesaian tugas akhir ini.
7. Pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu atas semua dukungannya baik secara langsung maupun tidak langsung.

Semoga Allah SWT membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Jakarta, 1 Agustus 2010

Penulis

## ABSTRACT

Text matching (string matching) is applied on most of information-based applications such as dictionary, information retrieval, and search engine. In general, string matching could be done in two approaches, exact string matching and approximate string matching.

One of approximate string matching method commonly implemented on database is the SoundEx algorithm, which is matching strings based on the sounds approximation of the text. Unfortunately, SoundEx algorithm is facing accuracy problems on text with a lot of articulating and spelling variations such as a name of person or place. The goal of this study is to analyze name-matching using SoundEx algorithm and using Jaro-Winkler algorithm in order to achieve better solution on application development with the feature of name-matching over database.

Scope of design and implementation of the software are limited to the use of name data samples, and excluding other data samples such as address and phone number. At the last phase of software development, name-matching results using both algorithms are evaluated. What has been done and what has not been done through this study will be reviewed in the last part of this thesis.

*Keywords: String Matching, SoundEx, Jaro-Winkler*

xv+63 pages; 21 figures; 8 tables; 1 attachments

Bibliography: 17 (1918-2010)



## ABSTRAK

Pencocokan teks (*string matching*) banyak dibutuhkan oleh aplikasi dalam dunia informatika. Contoh aplikasi yang melakukan *string matching* antara lain adalah kamus, *information retrieval*, dan mesin pencari. Proses pencocokan teks secara garis besar dilakukan dengan dua pendekatan, yaitu *exact string matching* (pencocokan secara tepat) dan *approximate string matching* (pencocokan berdasarkan kemiripan teks).

Salah satu metode *approximate string matching* yang umum diterapkan dalam basis data adalah dengan menggunakan algoritma SoundEx, yaitu pencocokan *string* berdasarkan kemiripan pengucapannya. Namun, algoritma SoundEx menghadapi masalah ketidakakuratan pada teks dengan banyak variasi ejaan dan pengucapan seperti pada nama orang atau nama tempat. Penelitian ini bertujuan untuk menganalisa tingkat kecocokan pencarian nama dengan menggunakan algoritma SoundEx dan juga algoritma Jaro-Winkler, dengan harapan hasilnya akan dapat dimanfaatkan sebagai solusi yang lebih baik dalam pembuatan aplikasi yang memanfaatkan pencarian nama dalam basis data.

Desain dan implementasi perangkat lunak dibatasi pada pencocokan terhadap sampel data nama saja, dan tidak melibatkan data pendukung yang lain seperti alamat dan nomor telepon. Pada tahap akhir pengembangan perangkat lunak, dilakukan analisa terhadap hasil kecocokan pencarian nama dengan menggunakan kedua algoritma tersebut di atas. Hal-hal apa yang telah dilakukan dan apa yang belum dilakukan pada penelitian ini akan diulas pada bagian akhir skripsi ini.

*Kata kunci: Pencocokan Kata, SoundEx, Jaro-Winkler*

xv+63 halaman; 21 gambar; 8 tabel; 1 lampiran

Daftar acuan: 17 (1918-2010)

## DAFTAR ISI

LEMBAR PERNYATAAN.....	iii
LEMBAR PERSETUJUAN .....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
ABSTRACT.....	viii
ABSTRAK.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL .....	xiv
DAFTAR LAMPIRAN .....	xv
BAB I PENDAHULUAN.....	1
1.1    LATAR BELAKANG.....	1
1.2    RUMUSAN MASALAH.....	2
1.3    TUJUAN PENELITIAN.....	4
1.4    RUANG LINGKUP .....	5
1.5    METODE PENELITIAN .....	6
1.5.1    Studi literatur.....	6
1.5.2    Perancangan .....	7
1.5.3    Implementasi.....	7
1.5.4    Pengujian dan analisa .....	8
1.5.5    Pengambilan kesimpulan .....	8
1.5.6    Penulisan laporan .....	8
1.6    SISTEMATIKA PENULISAN .....	9
BAB II LANDASAN TEORI.....	10
2.1    PENCOCOKAN <i>STRING</i> ( <i>STRING MATCHING</i> ) .....	10
2.1.1    Pengertian pencocokan <i>string</i> .....	10
2.1.2    Klasifikasi pencocokan <i>string</i> .....	11
2.1.3 <i>Phonetic string matching</i> .....	11
2.1.4 <i>Approximate string matching</i> .....	12

2.2	ALGORITMA SOUNDEX .....	14
2.2.1	Aturan kode fonetis .....	14
2.2.2	Penghitungan kode fonetis.....	16
2.3	ALGORITMA JARO-WINKLER .....	17
2.4	METODE PERANCANGAN.....	17
2.5	<i>UNIFIED MODELLING LANGUAGE</i> .....	18
2.5.1	Komponen uml.....	19
2.5.2	<i>Use case diagram</i> .....	20
2.5.3	<i>Activity diagram</i> .....	22
2.6	BASIS DATA .....	24
2.6.1	<i>Structured query language</i> .....	25
2.6.2	Microsoft sql server.....	26
2.6.3	Fungsi ( <i>user defined function</i> ) .....	27
2.6.4	Borland delphi.....	28
BAB III ANALISA DAN PERANCANGAN SISTEM .....		30
3.1	ANALISIS MASALAH .....	30
3.1.1	Algoritma soundex .....	30
3.1.1.1	Penerapan dalam Program.....	31
3.1.2	Algoritma jaro-winkler .....	33
3.1.2.1	Definisi jaro-winkler .....	33
3.1.2.2	Penerapan dalam Program.....	36
3.2	ANALISIS KEBUTUHAN .....	36
3.3	PERANCANGAN.....	37
3.3.1	<i>Use case diagram</i> .....	38
3.3.2	<i>Activity diagram</i> .....	39
3.3.3	<i>Activity diagram</i> pencocokan SoundEx.....	39
3.3.4	<i>Activity diagram</i> pencocokan Jaro-Winkler .....	41
3.3.5	Spesifikasi tabel .....	41
3.3.6	Rancangan antarmuka program.....	42
BAB IV IMPLEMENTASI DAN PENGUJIAN .....		43
4.1	KEBUTUHAN SISTEM .....	43
4.2	IMPLEMENTASI .....	44
4.2.1	Tampilan Program.....	44

4.3	PENGUJIAN.....	47
4.3.1	Pengumpulan Data Nama Sampel.....	48
4.3.2	Pengujian Fungsi Jaro-Winkler.....	49
4.3.3	Hasil Pencocokan Nama .....	52
4.3.3.1	Kemiripan Bunyi Nama .....	53
4.3.3.2	Kesalahan Ketik.....	56
4.3.3.3	Ketidakkelengkapan Nama .....	58
	BAB V KESIMPULAN.....	61
5.1	KESIMPULAN .....	61
5.2	SARAN.....	62
	DAFTAR PUSTAKA.....	64
	LAMPIRAN A <i>SOURCE CODE</i> .....	66

## DAFTAR GAMBAR

1. Gambar 2.1: Contoh use case diagram (Dharwiyanti dan Wahono 2003) .....	22
2. Gambar 2.2: Contoh activity diagram (Dharwiyanti dan Wahono 2003) .....	24
3. Gambar 3.1: Use case Cari Nama .....	38
4. Gambar 3.2: Activity diagram .....	39
5. Gambar 3.3: Activity diagram pencocokan SoundEx .....	40
6. Gambar 3.4: Activity diagram pencocokan Jaro-Winkler .....	41
7. Gambar 3.5: Rancangan antarmuka utama .....	42
8. Gambar 4.1: Antarmuka utama .....	45
9. Gambar 4.2: Antarmuka untuk mengelola data nama .....	45
10. Gambar 4.3: Antarmuka detail informasi pencocokan SoundEx .....	46
11. Gambar 4.4: Antarmuka detail informasi pencocokan Jaro-Winkler .....	46
12. Gambar 4.5: Hasil perbandingan nama 'FARID' dengan nama variasinya ...	49
13. Gambar 4.6: Hasil pencarian nama 'Devi Fitriana' .....	53
14. Gambar 4.7: Hasil pencarian nama 'Devi Fitriyana' .....	54
15. Gambar 4.8: Hasil pencarian nama 'Defi Fitriana' .....	54
16. Gambar 4.9: Hasil pencarian nama 'Dhevy Fitriana' .....	55
17. Gambar 4.10: Hasil pencarian nama 'Devi Firtiana' .....	56
18. Gambar 4.11: Hasil pencarian nama 'Dvei Fritiana' .....	57
19. Gambar 4.12: Hasil pencarian nama 'Devi' .....	58
20. Gambar 4.13: Hasil pencarian nama 'Dvei F.' .....	59
21. Gambar 4.14: Hasil pencarian nama 'D. Fitriana' .....	59

## DAFTAR TABEL

1. Tabel 2.1: Aturan pemberian kode pada algoritma SoundEx .....	14
2. Tabel 3.1: Hasil SoundEx beberapa variasi nama 'FARID' .....	31
3. Tabel 3.2: Skor kecocokan SoundEx menggunakan fungsi Difference .....	32
4. Tabel 3.3: Ilustrasi kecocokan teks .....	35
5. Tabel 3.4: Spesifikasi tabel Nama .....	42
6. Tabel 4.1: Ilustrasi kecocokan 'FARID' dan 'FARID' .....	49
7. Tabel 4.2: Ilustrasi kecocokan 'FARID' dan 'FARIDA' .....	50
8. Tabel 4.3: Ilustrasi kecocokan 'FARID' dan 'FAREED' .....	51

## DAFTAR LAMPIRAN

Lampiran A: SOURCE CODE .....	66
A.1: UDF JARO-WINKLER .....	66
A.2: SIMULASI.EXE .....	71
A.2.1: Main.dfm .....	71
A.2.2: Main.pas .....	79
A.2.3: DetailSoundex.pas .....	85
A.2.4: DetailJaroWinkler.pas .....	88