

IMPLEMENTASI ALGORITMA EDIT DISTANCE DALAM PENCOCOKAN STRING

TUGAS AKHIR



Oleh :

CAHYONO WIBOWO

22 05 3861

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

2012

IMPLEMENTASI ALGORITMA EDIT DISTANCE DALAM PENCOCOKAN STRING

TUGAS AKHIR



Diajukan kepada Fakultas Teknologi Informasi

Program Studi Teknik Informatika

Universitas Kristen Duta Wacana

Sebagai salah satu syarat dalam memperoleh gelar

Sarjana Komputer

Disusun oleh:

CAHYONO WIBOWO

22 05 3861

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

2012

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

IMPLEMENTASI ALGORITMA EDIT DISTANCE DALAM PENCOCOKAN STRING

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 keserjanaan 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 keserjanaan saya.

Yogyakarta, 22 Januari 2013



CAHYONO WIBOWO
22053861

HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI ALGORITMA EDIT DISTANCE
DALAM PENCOCOKAN STRING

Nama Mahasiswa : CAHYONO WIBOWO

N I M : 22053861

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2012/2013



Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 22 Januari 2013

Dosen Pembimbing I


Budi Susanto, SKom.,M.T.

Dosen Pembimbing II


Lukas Chrisantyo, M.Eng.

HALAMAN PENGESAHAN

IMPLEMENTASI ALGORITMA EDIT DISTANCE DALAM PENCOCOKAN STRING

Oleh: CAHYONO WIBOWO / 22053861

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 10 Januari 2013

Yogyakarta, 22 Januari 2013
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom., M.T.
2. Lukas Chrisantyo, M.Eng.
3. Junius Karel, M.T.
4. Rosa Delima, S.Kom., M.Kom.



Dekan


(Drs. Wimmie Handiwidjojo, M.T.)

Ketua Program Studi


(Nugroho Agus Haryono, M.Si)

HALAMAN PERSEMBAHAN

“Impossible is Nothing...”

©UKDW

Dengan penuh cinta dan kasih sayang

Kupersembahkan skripsi ini untuk

Tuhan Yesus, Alm. Papa, Mama, Cye-Cye, Koko, Ooh,

dan teman-temanku

Semoga bermanfaat bagi kita semua

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan anugerahNya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul IMPLEMENTASI ALGORITMA EDIT DISTANCE DALAM PENCOCOKAN STRING dengan baik dan tepat waktu.

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 dan mampu memberikan informasi yang berkualitas, 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 secara tidak langsung. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. Tuhan dan Juru Selamat pribadiku, Yesus Kristus atas pemeliharaan dan penyertaan-Mu dalam hidupku
2. Bapak Budi Susanto, S.Kom., M.T., selaku dosen pembimbing I yang telah memberikan bimbingannya dengan sabar dan baik kepada penulis, juga kepada
3. Bapak Lukas Chrisantyo, M.Eng., selaku dosen pembimbing II atas bimbingan, petunjuk dan masukan yang diberikan selama pengerjaan tugas ini sejak awal hingga akhir
4. Mama tercinta dan Keluarga tercinta yang selalu memberikan doa, dukungan, dan semangat
5. Teman seperjuangan Kharis Handoko dan teman-teman seperjuangan lainnya yang tidak dapat penulis sebutkan yang telah memberikan

inspirasi, masukan, dukungan dan semangat dalam proses pengerjaan skripsi ini

6. 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. Semoga karya ini dapat berguna bagi kita semua.

Yogyakarta, 22 Januari 2013

Penulis

©UKYDWN

INTISARI

Structured Query Language atau sering dikenal dengan istilah *SQL* merupakan bahasa standar yang digunakan untuk mengolah data-data pada database. Pada umumnya sistem pencarian pada *SQL* menggunakan perintah “*like*”, yang berarti sistem akan melakukan pencarian data sesuai dengan masukan dari pengguna sistem. Hal ini tentu tidaklah efektif apabila pengguna sistem melakukan kesalahan pengetikan dalam melakukan masukan string. Hasil yang ditampilkan oleh sistem tidak sesuai dengan yang diinginkan oleh pengguna sistem.

Sistem ini akan mencocokkan kata yang berupa string dari masukan pengguna dengan string yang ada dalam database. Dengan masukan dari user yang memungkinkan kesalahan pengetikan untuk string *input* / masukan, sistem akan mengelompokkan data yang ada dalam database berdasarkan jumlah huruf dari string masukan. Kemudian dari hasil pengelompokan data tersebut, sistem akan membandingkan string tersebut dengan string masukan dari pengguna dengan metode *Edit Distance*. Hasil dari algoritma *Edit Distance* berupa variabel k yang akan menentukan string tersebut layak atau tidak masuk dalam pengelompokan daftar saran.

Kesimpulan yang diperoleh dari penelitian ini adalah program aplikasi yang dibangun dengan mengimplementasikan algoritma *Edit Distance*. Sistem pencocokan string dengan algoritma *Edit Distance* $k = 0,1,2$ dan batasan jumlah karakter ± 1 , menghasilkan jumlah saran yang memiliki kemiripan kata dengan kata kunci pencarian lebih besar dan tidak menghasilkan jumlah saran yang terlalu banyak.

Kata Kunci : algoritma *Edit Distance*, pencocokan *string*

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN	v
HALAMAN PERSEMBAHAN	vi
UCAPAN TERIMA KASIH	vii
INTISARI	ix
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	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	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.1.1 String	5
2.1.2 String Matching	6

2.2	Landasan Teori	8
2.2.1	Algoritma Edit Distance	8
2.2.2	Contoh Algoritma Edit Distance	9
BAB 3 PERANCANGAN SISTEM		14
3.1	Analisis Kebutuhan Sistem	14
3.1.1	Perangkat Keras (<i>Hardware</i>)	14
3.1.2	Perangkat Lunak (<i>Software</i>)	14
3.2	Perancangan Sistem	15
3.2.1	Arsitektur Sistem	15
3.2.2	Algoritma Edit Distance	16
3.2.3	Algoritma Pembuatan Program	18
3.3	Perancangan Database	19
3.4	Perancangan Antarmuka Program	21
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM		28
4.1.1	Antarmuka Program	28
4.1.1.1	User Login	28
4.1.1.2	Admin	29
4.1.1.2	Pencarian Buku	29
4.1.1.3	Keterangan	31
4.1.1.4	Pembuatan Id Baru	31
4.1.1.5	Lupa Kata Kunci	32
4.1.1.6	Penambahan Data	33
4.1.1.7	Pengubahan Data	34
4.1.1.8	Penghapusan Data	34
4.1.2	Implementasi Algoritma Edit Distance	35

4.2 Analisis Sistem	36
4.2.1 Analisis Algoritma Edit Distance	36
4.2.2 Analisis Program	38
4.3 Kelebihan dan Kekurangan Sistem	39
BAB 5 KESIMPULAN DAN SARAN	41
5.1 Kesimpulan	41
5.2 Saran	41
DAFTAR PUSTAKA	
LAMPIRAN	

©UKDW

DAFTAR TABEL

Tabel 3.1 Database User	19
Tabel 3.2 Database Siswa	20
Tabel 3.3 Database Wali Siswa.....	20

©UKDW

DAFTAR GAMBAR

Gambar 2.1 Algoritma Computing Edi Distance	9
Gambar 2.2 Langkah 1- 5.....	11
Gambar 2.3 Langkah 6 – 12 ketika $i = 1$	11
Gambar 2.4 Langkah 6 – 12 ketika $i = 2$	12
Gambar 2.5 Langkah 6 – 12 ketika $i = 3$	12
Gambar 2.6 Langkah 6 – 12 ketika $i = 4$	13
Gambar 2.7 Langkah 6 – 12 ketika $i = 5$	13
Gambar 3.1 Use case Sistem	14
Gambar 3.2 Flowchart Algoritma Edit Distance	17
Gambar 3.3 Flowchart Alur Program	19
Gambar 3.4 Relasi Database	21
Gambar 3.5 Desain masuk pengguna.....	21
Gambar 3.6 Desain Pencarian data	22
Gambar 3.7 Desain data siswa dan wali siswa.....	23
Gambar 3.8 Desain menu utama admin	24
Gambar 3.9 Desain membuat id baru.....	24
Gambar 3.10 Desain lupa kata kunci	25
Gambar 3.11 Desain tambah data	25
Gambar 3.12 Desain ubah data	26
Gambar 3.13 Desain hapus data.....	27
Gambar 4.1 Halaman Masuk.....	28
Gambar 4.2 Halaman Menu Utama Admin	29

Gambar 4.3 Halaman Pencarian.....	30
Gambar 4.4 Halaman Keterangan.....	31
Gambar 4.5 Halaman Pembuatan Id Baru.....	32
Gambar 4.6 Halaman Lupa Kata Kunci.....	33
Gambar 4.7 Halaman Penambahan data.....	33
Gambar 4.8 Halaman Pengubahan data.....	34
Gambar 4.9 Halaman Penghapusan data.....	35
Gambar 4.10 Analisis Edit Distance.....	37
Gambar 4.11 Analisis Program.....	39

©UKDW

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Sistem pencarian data dalam dunia komputer menjadi sesuatu yang berguna dalam mengolah data. Dengan melakukan pencarian yang baik dan benar, manusia dapat mengolah data-data komputer dengan waktu yang relatif cepat dan efisien. Apabila data pencarian tidak diisi dengan benar, maka hasil / output dari pencarian data akan menyimpang dari data yang seharusnya diolah. Hal ini sering kali terjadi dalam mengolah data yang beribu-ribu banyaknya. Akibatnya banyak waktu yang terbuang untuk mencari data yang ingin diolah dan membuat tingkat *stress* manusia meningkat.

Structured Query Language atau sering dikenal dengan istilah SQL merupakan bahasa standar yang digunakan untuk mengolah data-data pada database. Dengan menggunakan SQL pengolahan data-data pada database server menjadi lebih mudah.

Pada umumnya sistem pencarian pada SQL menggunakan perintah "like", yang berarti sistem akan melakukan pencarian data sesuai dengan masukan dari pengguna sistem. Hal ini tentu tidaklah efektif apabila pengguna sistem melakukan kesalahan pengetikan dalam melakukan masukkan string. Hasil yang ditampilkan oleh sistem tidak sesuai dengan yang diinginkan oleh pengguna sistem.

Berdasar latar belakang diatas, penulis akan membangun sistem dengan menerapkan algoritma Edit Distance untuk membantu user dalam melakukan pencarian data. Dengan masukan dari user yang memungkinkan kesalahan pengetikan untuk string input, sistem dapat memberikan saran untuk string yang memungkinkan penulisan string yang lebih benar untuk menghemat waktu pencarian data yang akan diolah.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah diatas, maka penulis akan membangun sebuah sistem dengan teori dari *Edit Distance* untuk mencocokkan input *string* pencarian data dengan data-data dari database *string* pencarian yang mempunyai kemiripan kata. Teori dari *Edit Distance* digunakan dalam mencari beberapa kemungkinan *string* yang mirip dengan input pencarian.

Masalah yang akan diteliti adalah: Menerapkan *Edit Distance* untuk dapat merangking kemiripan *string* berdasarkan pola *string* yang dimasukkan.

1.3 Batasan Masalah

Batasan masalah pada pengerjaan tugas akhir ini adalah sebagai berikut:

- Variabel k perbedaan yang akan digunakan di dalam sistem pencocokan string adalah sampai pada level 2.
- Sistem pencocokan string hanya mencocokkan data-data siswa dan wali siswa.
- Sistem pencarian data berdasar nama siswa dan wali.
- Hasil dari pencocokan string dengan masukan akan ditampilkan secara urut dimulai dengan perbedaan nilai $k = 0$.
- Sistem yang akan dibangun berupa aplikasi program desktop daftar siswa yang hanya menyediakan data-data siswa dan wali siswa.

1.4 Tujuan Penelitian

Tujuan utama:

- Mendapatkan data hasil dari input pencarian data dengan benar dan cepat

Sub tujuan:

- Membangun sebuah aplikasi program dekstop yang mampu menyediakan data-data pencocokan *string* sebagai pembanding dengan input *string* dari sistem pencarian data, serta mudah digunakan oleh pengguna sistem.
- Menghasilkan output sistem yang mampu memberikan alternatif luaran dalam input pencarian. Sehingga memudahkan pengguna dalam mencari data-data yang diinginkan.

1.5 Metode Penelitian

Metode atau pendekatan yang akan digunakan untuk merealisasikan tujuan dan pemecahan masalah diatas adalah:

- Studi Pustaka (literature)
Studi pustaka dilakukan dengan membaca sumber-sumber pustaka, berupa buku-buku serta sumber-sumber online di internet yang dapat dipercaya, seperti jurnal-jurnal internasional dan makalah-makalah ilmiah. Tujuannya adalah untuk mengumpulkan data-data dalam pembuatan sistem ini.
- Studi Lapangan, melalui penelitian lapangan, yaitu dengan:
 1. Melakukan pengumpulan data melalui wawancara dengan pihak sekolah untuk memperoleh keterangan dan data yang dibutuhkan untuk mendukung penyelesaian Tugas Akhir.
 2. Melakukan konsultasi dengan dosen pembimbing berhubungan dengan perancangan dan pelaporan tugas akhir.
 3. Melakukan pengujian sistem dan melakukan perbaikan apabila menemukan kesalahan-kesalahan.

1.6 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir dibagi dalam beberapa bab yang masing-masing memiliki penjelasan yaitu:

BAB 1 berisi PENDAHULUAN yang mencakup tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan laporan.

BAB 2 berisi TINJAUAN PUSTAKA yang mencakup penjelasan tentang tinjauan pustaka dan teori-teori dasar yang diperlukan untuk pembuatan tugas akhir ini.

BAB 3 berisi PERANCANGAN SISTEM yang mencakup rancangan kerja sistem, rancangan antar muka aplikasi yang akan dibangun, serta aliran data dalam aplikasi yang akan dibuat.

Bab 4 berisi IMPLEMENTASI dan ANALISIS SISTEM yang membahas tentang implementasi perancangan yang telah dibuat pada BAB 3 ke dalam bahasa pemrograman.

Bab 5 berisi PENUTUP yang mencakup kesimpulan tentang sistem yang dihasilkan, serta beberapa saran yang dapat digunakan untuk perbaikan dan pengembangan sistem pada masa mendatang.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan oleh penulis, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Faktor-faktor yang mempengaruhi nilai k dalam pencocokan string adalah pemindahan (transposisi), penambahan, penghapusan, dan pergantian karakter (substitusi)
2. Algoritma edit distance yang telah dikembangkan mampu memberikan hasil yang lebih maksimal dalam hal memperhitungkan jarak kesalahan perubahan string. Hal ini berarti saran-saran kata yang dihasilkan dapat lebih akurat
3. Jarak kesalahan 2 yang digunakan merupakan jarak kesalahan paling maksimal karena akan menghasilkan saran kata yang tidak menyimpang terlalu jauh.
4. Proses pencarian berdasar tokensasi yang digunakan dalam proses pencarian kata membantu memperakurat hasil data dan mempersingkat waktu pencarian.

5.2 Saran

Sebagai saran penelitian berikutnya adalah :

1. Perlunya dikembangkan algoritma Edit Distance agar sistem semakin dinamis dalam menghasilkan kemiripan data tertentu
2. Perlunya dilakukan pengelompokan data lebih lanjut agar sistem semakin akurat dan efisien dalam menentukan kemiripan data
3. Menambahkan fitur daftar saran yang sering dilihat oleh pengguna sistem sebelumnya

4. Algoritma edit distance dapat juga diimplementasikan untuk permasalahan lain seperti identifikasi kemiripan DNA, pencarian string dengan tingkat presentase tertentu, dan perbandingan tingkat kemiripan antara 2 dokumen.
5. Sistem dapat dikembangkan lagi sehingga dapat dijalankan berbasis web.

©UKDW

DAFTAR PUSTAKA

- Cheriet, M., Kharma, N., Cheng-Lin Liu, Ching Suen. 2007. *Character Recognition Systems : A Guide for Students and Practitioners*. United States of America; John Wiley & Sons, Inc.
- Kusumo, A.S. 2006. *Pemrograman Visual Basic 2005*. Jakarta; PT Elex Media Komputindo.
- Jurafsky, D. & Martin, JH. 2000. *Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey; Prentice Hall.
- Navarro, G. 2001. *A Guided Tour to Approximate String Matching*. University of Chile ; ACM Computing Surveys Vol 33 (1), 31-88.
- Nugues, PM. 2006. *An Introduction to Language Processing with Perl and Prolog*. Berlin; Springer.
- Santoso, H. 2005. *Membangun Aplikasi .NET menggunakan VB.Net 2005*. Jakarta; PT Elex Media Komputindo.
- Solution, W. 2009. *Sistem Informasi Manufaktur dengan VB 2005 dan SQL Server 2005*. Jakarta; PT Elex Media Komputindo.
- Syaroni, M. & Munir, R. 2005. *Pencocokan String Berdasarkan Kemiripan Ucapan (Phonetic String Matching) dalam Bahasa Inggris*. SNATI 18 Juni 2005. Yogyakarta.
- Gilleland, Michael. 2000. *Levenshtein Distance in Three Flavors*. dalam <http://www.merriam.com/id>; 4 Desember 2012.