

## PERBANDINGAN PENCOCOKAN STRING NAMA DALAM BAHASA INDONESIA BERDASARKAN KEMIRIPAN UCAPAN (PHONETIC STRING MATCHING) SOUNDEX DAN DOUBLETAPHONE

Sukma Rahadian<sup>1</sup>, Rimba Whidiana<sup>2</sup>, Adiwijaya<sup>3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

### Abstrak

Ketika akan mencocokkan beberapa buah string ada baiknya memilih algoritma yang mudah dan sederhana namun menghasilkan solusi yang optimal dan bisa menjamin kepuasan si pengguna. Beberapa string yang dibandingkan bisa dikatakan sama atau mirip jika mempunyai kesamaan atau kemiripan, apakah itu sama berdasarkan kemiripan dalam hal penulisan atau sama berdasarkan kemiripan dalam hal pengucapan. Dalam pencocokan string berdasarkan kemiripan ucapan ada beberapa algoritma yang bisa digunakan. Algoritma yang bisa digunakan adalah algoritma Soundex dan algoritma Doublemetaphone. Hanya saja kedua algoritma ini tidak bisa menghasilkan solusi yang optimal jika digunakan dalam mencocokkan string nama orang dalam ejaan bahasa Indonesia, karena ejaan yang digunakan dalam kedua algoritma ini adalah ejaan dalam bahasa Inggris. Selain terdapat perbedaan dalam ejaan, hal yang menjadi kendala adalah adanya ejaan lama dalam bahasa Indonesia yang masih diterapkan dalam penamaan orang Indonesia. Hal tersebut menjadi kerugian bagi kedua algoritma ini karena algoritma Soundex dan algoritma Doublemetaphone tidak bisa mendeteksi ejaan lama tersebut. Sehingga dengan demikian solusi yang bisa diambil adalah dengan melakukan perubahan ejaan pada algoritma Soundex dan algoritma Doublemetaphone agar kedua algoritma ini bisa mendeteksi string nama orang dalam ejaan bahasa Indonesia. Kemudian ditambahkan algoritma tambahan untuk mendeteksi ejaan lama dalam bahasa Indonesia yang disebut sebagai tahap Normalisasi.

Kata Kunci : perbandingan string, soundex, doublemetaphone, ejaan, normalisasi.

### Abstract

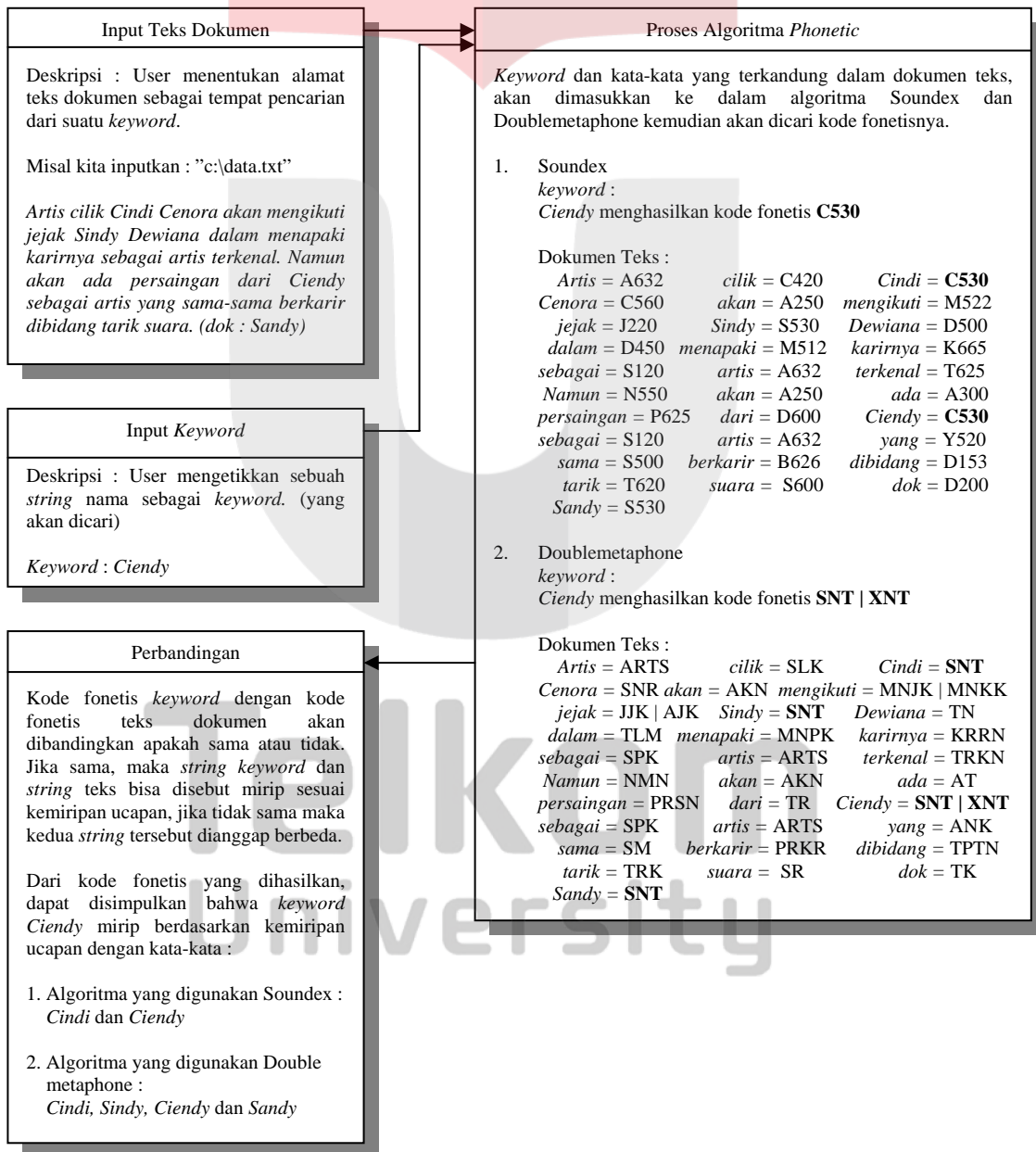
On matching several strings, it is better to choose an easy and simple algorithm that generates the optimal solution and ensures the satisfaction of the users. Strings can be considered equal or alike if they have similarity, whether it is based on the similarity in writing or pronunciation. When we match strings based on the pronunciation, there are several algorithms that can be used. They are Soundex algorithm and Doublemetaphone algorithm. However, these two algorithm cannot produce optimal solution when comparing strings of people's names in Bahasa Indonesia. This is because the algorithm presented uses the English spelling method. Other than that, the other problem on using these algorithm is the use of old spelling that is still applied on names of Indonesian people. These become the drawback of the algorithms, cause they cannot detect those old spelling name. To overcome those weaknesses, a solution is proposed to change the spelling method used on the Soundex and Doublemetaphone algorithm so that they can detect strings of people's name in Bahasa Indonesia. Moreover, an additional algorithm is used to detect old spelling in Bahasa Indonesia called as the Normalization step. As a result a modified algorithms will be constructed as to enhanced the functionality of the algorithm.

Keywords : string matching, soundex, doublemetaphone, spelling, normalization.

# 1. Pendahuluan

## 1.1 Latar belakang

Pencocokan *string* akan sangat membantu jika menggunakan algoritma yang sesuai, saat ini terdapat banyak algoritma yang bisa digunakan. Beberapa yang bisa digunakan adalah algoritma Soundex dan algoritma DoubleMetaphone pada teknik *Phonetic String Matching*.



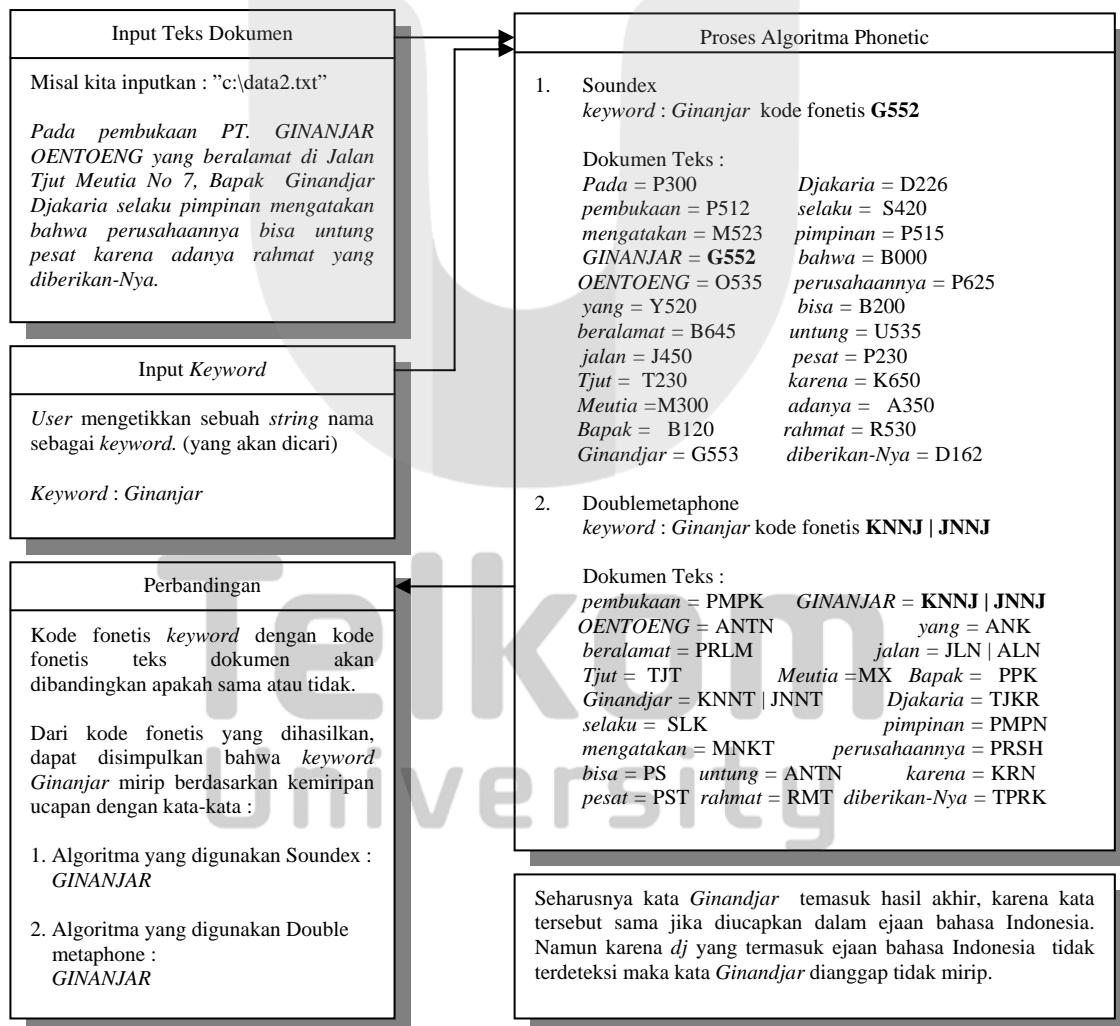
Gambar 1-1: Bagan Sistem Phonetic String Matching(a)

*Phonetic String Matching* merupakan teknik mencocokkan *string* berdasarkan kemiripan ucapan. Sebuah *string* dikatakan mirip jika mempunyai kode fonetis yang sama. Namun yang jadi permasalahan adalah bagaimana algoritma tersebut bisa mendeteksi nama orang Indonesia karena nama Indonesia dan nama Inggris ada beberapa perbedaan dalam kode fonetisnya.

Algoritma yang ada sekarang ini diciptakan berdasarkan ejaan bahasa Inggris. Sehingga jika digunakan untuk mencocokkan *string* nama dalam pengucapan bahasa Indonesia akan kurang maksimal. Oleh karena itu diperlukan modifikasi kedua algoritma tersebut diatas agar bisa bekerja lebih baik.

## 1.2 Perumusan masalah

Permasalahan yang muncul dari algoritma Soundex dan Doublemetaphone yang ada sekarang ini adalah kurang maksimalnya algoritma dalam mendeteksi ejaan Indonesia, karena ejaan yang digunakan adalah ejaan Inggris. Salah satu contohnya diperlihatkan melalui Gambar 1-2.



Gambar 1-2: Bagan Sistem Phonetic String Matching(b)

Permasalahan yang dijadikan objek penelitian tugas akhir ini adalah :

1. Algoritma Soundex dan algoritma DoubleMetaphone masih menggunakan ejaan bahasa Inggris, kurang maksimal jika digunakan untuk mendeteksi *string* nama orang Indonesia. Sehingga perlu dilakukan modifikasi.
2. Algoritma Soundex dan algoritma Doublemetaphone tidak bisa mendeteksi beberapa ejaan lama bahasa Indonesia.
3. Menentukan algoritma mana yang baik digunakan untuk mendeteksi nama orang Indonesia berdasarkan keakuratan, kesensitifan algoritma terhadap tanda petik, dan terhadap panjangnya *string* nama yang dimasukkan.

Batasan masalah yang diberikan sebagai berikut :

1. Teks dokumen yang dijadikan sebagai sumber pencarian hanya teks dokumen dengan file yang berakhiran .txt.
2. Karakter khusus yang diperbolehkan dalam *string* nama hanya tanda petik.

### 1.3 Tujuan

Tujuan dari penelitian tugas akhir ini adalah :

1. Memodifikasi algoritma Soundex dan algoritma DoubleMetaphone menjadi algoritma *phonetic* yang ber-ejaan bahasa Indonesia disertai penanganan terhadap ejaan lama bahasa Indonesia.
2. Membandingkan algoritma Soundex dengan algoritma Doublemetaphone, dalam melakukan pencocokan *string* nama orang Indonesia yang mencakup keakuratan, kesensitifan algoritma terhadap tanda petik, dan terhadap panjangnya *string* nama yang dimasukkan.

### 1.4 Metodologi penyelesaian masalah

Metodologi penyelesaian masalah yang digunakan dalam penelitian tugas akhir ini adalah :

1. Studi literatur
  - a. Pencarian dan pengumpulan kajian-kajian yang berkaitan dengan masalah-masalah yang ada pada tugas akhir ini, baik berupa artikel, buku referensi, mencari referensi via internet dan sumber-sumber lain yang berhubungan dengan masalah tugas akhir. Termasuk melakukan bimbingan dengan dosen pembimbing.
  - b. Pengumpulan data-data pendukung dan melakukan analisa (spesifikasi) sistem yang diperlukan dalam tahap implementasi.
2. Analisis perancangan dan implementasi  
Perancangan dan implementasi aplikasi perangkat lunak sesuai analisa sistem.
3. Uji coba terhadap aplikasi perangkat lunak yang telah dibuat dan menganalisa performansi dari algoritma yang digunakan.
4. Pembuatan laporan tugas akhir.

## 5. Kesimpulan dan Saran

### 5.1 Kesimpulan

Kesimpulan yang dapat diambil dari Tugas Akhir “Perbandingan Pencocokan String Nama Dalam Bahasa Indonesia Berdasarkan Kemiripan Ucapan (Phonetic String Matching) Soundex dan Doublemetaphone” adalah :

1. Modifikasi Algoritma Soundex dan Doublemetaphone mampu mendeteksi *string* nama ber-ejaan bahasa Indonesia dan hasil deteksinya memiliki tingkat akurasi yang lebih baik dibandingkan algoritma Soundex dan Doublemetaphone original.
2. Algoritma Soundex bisa menjadi pilihan yang baik untuk algoritma *phonetic string matching* dalam mendeteksi string nama ber-ejaan bahasa Indonesia dibandingkan algoritma Doublemetaphone modifikasi.

### 5.2 Saran

Saran untuk Tugas Akhir “Perbandingan Pencocokan String Nama Dalam Bahasa Indonesia Berdasarkan Kemiripan Ucapan (Phonetic String Matching) Soundex dan Doublemetaphone” adalah :

1. Algoritma Doublemetaphone modifikasi yang mengandung ejaan Indonesia digabungkan dengan Doublemetaphone original, sehingga menambah keakuratan dan kekuatan Doublemetaphone dalam mendeteksi kemungkinan banyak *string* nama yang bervariasi dalam berbagai bahasa.
2. Algoritma Soundex dan Doublemetaphone bisa diimplementasikan dalam internet sebagai *Search Engine* dengan teknik pencarian *phonetic string matching*.

## Daftar Pustaka

- [1] Badudu, Yus, 1990, "Ejaan Bahasa Indonesia". Pustaka Prima Bandung.
- [2] Carstensen, Adam, "An Introduction To Double Metaphone And The Principles Behind Soundex", <http://www.datamanagementgroup.com/resources/Articles/>, Diakses pada tanggal 22 Januari 2007 pukul 22.00 WIB
- [3] Depdikbud, 1988, "Tata Bahasa Baku Bahasa Indonesia". Balai Pustaka Jakarta.
- [4] Double Metaphone From wikipedia, the free encyclopedia, [http://en.wikipedia.org/wiki/Double\\_Metaphone](http://en.wikipedia.org/wiki/Double_Metaphone), Diakses pada tanggal 22 Januari 2007 pukul 22.00 WIB.
- [5] Karhendana, Arie dkk, 2003, "Normalsasi *String* untuk Optimasi Phonetic *String* Matching dalam Bahasa Indonesia". Departemen Teknik Informatika ITB Bandung.
- [6] Modul Implementasi Algoritma DoubleMetaphone. <http://www.cpan.org/modules/author/id/MSCHWEN>, Diakses pada tanggal 19 Januari 2007 pukul 23.30 WIB
- [7] Modul PHP Manual, <http://www.php.net/soundex>, Diakses pada tanggal 19 Januari 2007 pukul 23.30 WIB.
- [8] Munawaroh, 2004, "Nama Nama Populer dalam 24 Bahasa". Eska Media Jakarta.
- [9] Nurcahyo, Priyadi Iman, "Algoritma Fonetik Bahasa Indonesia", <http://priyadi.net/archives/2005/12/21/algoritma-fonetik-bahasa-ndonesia/>, Diakses pada tanggal 22 Januari 2007 pukul 22.00 WIB
- [10] Poerwadarminta, W.J.S, 2003, "Kamus Umum Bahasa Indonesia Edisi Ketiga". Balai Pustaka Jakarta.
- [11] Sidiq, Betha, 2006, "Pemrograman Web dengan PHP", Informatika Bandung.
- [12] Soundex From wikipedia, the free encyclopedia, <http://en.wikipedia.org/wiki/Soundex>, Diakses pada tanggal 22 Januari 2007 pukul 22.00 WIB.
- [13] Syafii, M, 2004, "Membangun Aplikasi Berbasis PHP dan MySQL". Penerbit Andi Yogyakarta.
- [14] Syaroni, Mukhamad dkk, 2004, "Pencocokan *String* Berdasarkan Kemiripan Ucapan (Phonetic *String* Matching) dalam Bahasa Inggris". Departemen Teknik Informatika ITB Bandung.
- [15] The Soundex Algorithm, [http://www.comp.leeds.ac.uk/matthewb/ar32/basic\\_soundex.htm](http://www.comp.leeds.ac.uk/matthewb/ar32/basic_soundex.htm), Diakses pada tanggal 22 Januari 2007 pukul 22.00 WIB