

Implementasi Rancang Sistem Pencarian Kata Pada Kamus Bahasa Indonesia Menggunakan Metode Approximate String Matching Berbasis Web

Devin Febrian Kurnia^{1*}, Hidayatullah Al Islami¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}devin.febrian.kurnia@gmail.com, ²dosen02408@unpam.ac.id

(* : coresponding author)

Abstrak—Kamus merupakan sebuah rujukan yang menerangkan makna atau arti dari sebuah kata. Penggunaan kamus di era modern saat ini berbeda dengan kamus konvensional yang memiliki tebal dan bobot yang sangat berat. Mencari kata menggunakan kamus konvensional yang juga memiliki kendala seperti memakan waktu dikarenakan hurufnya yang kecil. Jenis kamus pun beragam dengan semakin berkembangnya teknologi, salah satunya kamus online. Namun kendala yang pasti dialami adalah sangatnya berdampak pada informasi yang didapatkan oleh pengguna, salah satunya tidak mendapatkan kata yang tepat atau sama sekali tidak mendapatkan kata yang dicari. Dengan demikian diperlukan suatu sistem pencarian kata pada kamus bahasa Indonesia yang mampu memberikan saran kata yang diketikkan oleh pengguna. Salah satu metode dalam pengoreksian ejaan kata adalah *Approximate String Matching*, metode ini menerapkan pendekatan pencarian string. *Algoritma Levenshtein Distance* merupakan salah satu bagian metode *Approximate String Matching*. *Algoritma Levenshtein Distance* memiliki 3 macam operasi *string* yaitu penghapusan, penambahan dan pengubahan.

Kata Kunci: *Approximate String Matching*, Kata, Kamus, Bahasa Indonesia.

Abstract—*The dictionary is a reference that explains the meaning or meaning of a word. The use of dictionaries in the modern era is different from conventional dictionaries which have thick and very heavy weights. Finding words using conventional dictionaries also has constraints such as time consuming and small letters. The types of dictionaries also vary with the development of technology, one of which is an online dictionary. But the obstacle that must be experienced is not knowing for sure the word or finger slip that results in typo. This greatly affects the information obtained by the user, one of which is not getting the right words or not getting the words you are looking for. Thus we need a word search system in the Indonesian dictionary that is able to provide word suggestions typed by the user. One method of correcting word spelling is Approximate String Matching, this method applies a string search approach. The Levenshtein Distance Algorithm is a part of the Approximate String Matching method. The Levenshtein Distance algorithm has 3 kinds of string operations, namely deletion, addition and change.*

Keywords: *Approximate String Matching, Words, Dictionary, Indonesian*

1. PENDAHULUAN

Bahasa Indonesia amat penting peranannya untuk kehidupan bermasyarakat, berbangsa dan bernegara. Terlebih lagi bahasa Indonesia pun memiliki fungsi yang amat penting pada lingkungan media pendidikan dan media pengajaran. Menggunakan bahasa Indonesia yang baku yang baik dan benar bertujuan guna memabarkan kemampuan menggunakan bahasa Indonesia yang baku dengan segala kebutuhannya, yaitu sebagai media komunikasi, alat berpikir, media kebudayaan dan persatuan. Penggunaan bahasa Indonesia yang baku juga dapat digunakan untuk beberapa faktor keterampilan seperti media untuk menyimak, media untuk berbicara, media untuk membaca dan media untuk menulis. Keempat keterampilan berkaitan antara lain dengan cara yang bervariasi.

Media tulis adalah kemampuan seorang untuk mengungkapkan ide, pikiran, pengetahuan, ilmu dan pengalaman – pengalaman hidupnya dalam bahasa tulis yang baku, runtut, enak di baca, dan dipahami oleh orang lain. Menulis dapat dilakukan dimana saja dan melalui media apa saja. Naskah tulisan pada media cetak seperti buku, koran, atau majalah akan diedit oleh editor terlebih dahulu sebelum diterbitkan. Penggunaan kalimat baku yang sesuai dengan kaidah Bahasa Indonesia tidak hanya berlaku di bidang jurnalistik, namun berlaku juga pada penulisan naskah skripsi maupun *paper*.

Terdapat ajaran penulisan yang harus diperhatikan seperti penggunaan kata yang baku dan kata yang tidak baku. Khususnya untuk beberapa hal yang berkaitan dengan dokumen ilmiah seperti laporan, makalah, skripsi, karya ilmiah dan dokumen – dokumen resmi yang harus menyertai kalimat baku sesuai dengan EYD yang telah diberikan. Kesalahan pengetikan yang sering terjadi dapat menyebabkan kata baku berubah menjadi kata tidak baku karena pengejaan kata yang digunakan tidak sesuai dengan EYD.

Kesalahan pengejaan pada editor kata sering kali terjadi. Pengecekan kesalahan pengetikan biasa dilakukan ketika karya ilmiah telah selesai dibuat. Untuk kalimat yang pendek, hal itu kemungkinan tidak akan sulit untuk dilakukan. Tetapi, apabila ukuran kata sudah mencapai lebih dari puluhan, ribuan, ratusan bahkan jutaan kata, Pemeriksaan dengan cara tersebut sangat menyulitkan.

Approximate String Matching bisa digunakan untuk pencarian kata string berdasarkan kata *string* yang sama dan *string* yang memiliki kemiripan penulisan dengan *string* yang terdapat pada kamus. Metode ini dapat digunakan untuk pencarian kata tidak baku karena dapat mengidentifikasi *string* yang sama dan yang memiliki kemiripan penulisan.

Permasalahan yang muncul karena masih terdapat banyaknya pengguna individu yang masih kesulitan untuk mencari kata baku dalam kamus fisik sehingga dapat memakan banyak waktu untuk mencarinya dan masih banyaknya individu yang belum dapat membedakan kata dalam kamus bahasa Indonesia sehingga sulit untuk membedakan mana kata baku dan non baku. (Adiwidya, 2019)

Dengan dibuatnya implementasi rancang sistem pencarian kata pada kamus besar bahasa Indonesia menggunakan metode *approximate string matching* berbasis *web* ini yang bertujuan untuk membantu meringankan individual dalam proses pengecekan kesalahan penulisan kata pada sebuah naskah dan mengubahnya menjadi kata yang benar, serta membantu struktur kalimat baku dalam kamus besar bahasa Indonesia.

2. METODOLOGI PENELITIAN

2.1 *Approximate String Matching*

Approximate string matching merupakan metode yang digunakan untuk melakukan pencocokan atau kedekatan string dengan dasar kemiripan dari segi penulisannya dengan tingkat kemiripan ditentukan dengan jauh tidaknya beda penulisan dua buah string yang dibandingkan tersebut. (Friendly, 2018) Operasi sederhana yang dilakukan terdiri atas penyisipan, penghapusan, substitusi, dan transposisi. Tiap-tiap algoritma memiliki ciri khas operasi atau proses masing-masing.

2.2 *Levenshtein Distance*

Levenshtein Distance merupakan suatu pengukuran untuk menghitung jumlah perbedaan antara dua kata (Ariyani, 2016). Penghitungan jarak antara dua kata ditentukan dari jumlah minimum operasi perubahan untuk mengubah kata A menjadi kata B. Operasi yang dapat dilakukan oleh algoritma ini yaitu:

a. Operasi Pengubah Karakter

Operasi menukar sebuah karakter dengan karakter lain contohnya : “aju” menjadi “aku”. Karakter ‘j’ diganti dengan ‘k’.

b. Operasi Penambahan Karakter

Operasi menambah karakter kedalam sebuah kata contohnya : “brikut” menjadi “berikut”. Penambahan karakter tidak bisa lakukan diawal, diakhir atau disisipkan dalam kata.

c. Operasi Penghapusan Karakter

Operasi penghapusan karakter dilakukan dengan menghilangkan sebuah karakter dalam kata contohnya : “sampaio” menjadi “sampai” dengan menghilangkan karakter terakhir ‘o’.

3. ANALISA DAN PEMBAHASAN

3.1 Analisa Sistem

Analisis Sistem (System Analysis) merupakan suatu teknik atau metode pemecahan masalah dengan cara menguraikan sistem ke dalam komponen untuk mengetahui bagaimana komponen-komponen tersebut bekerja dan saling berinteraksi satu sama lain untuk mencapai tujuan sistem. Penulis akan menganalisa sistem dengan membuat sebuah rancangan sistem. Rancangan sistem yakni salah satu langkah dalam teknik pemecahan masalah di mana komponen-komponen pembentuk sistem digabungkan sehingga membentuk satu kesatuan sistem yang utuh. Teknik dari rancangan sistem ini meliputi proses penambahan, penghilangan, dan perubahan komponen-komponen dari sistem semula.

3.1.1 Analisa Sistem Berjalan

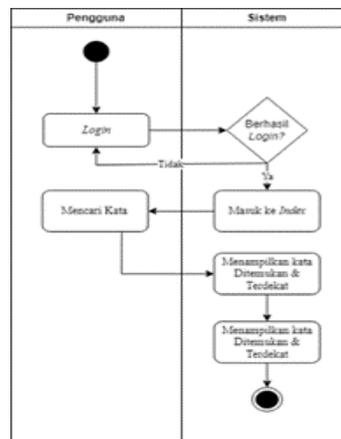
Adapun hasil analisa yang dilakukan tentang sistem yang sedang berjalan adalah masih banyaknya sekelompok atau individu yang masih menggunakan kamus fisik sebagai sarana pencarian kalimat baku bahasa Indonesia. Dengan melakukan pencarian manual tersebut akan membutuhkan banyak waktu karena pengguna harus mencari kalimat yang banyak pada satu buku kamus besar bahasa Indonesia.



Gambar 1. Analisa Sistem Berjalan

3.1.1 Analisa Sistem Usulan

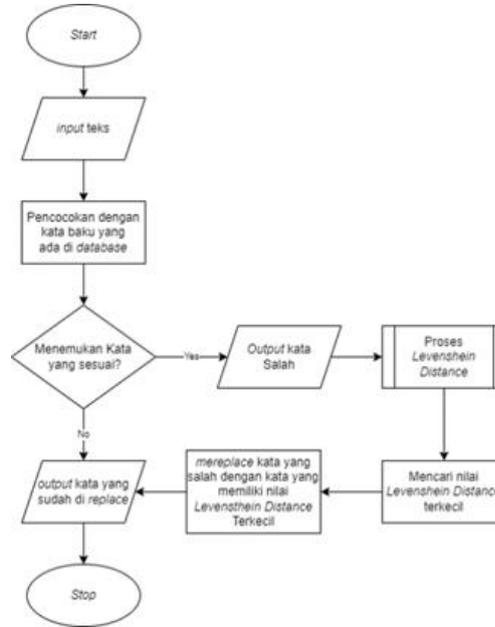
Bedasarkan sistem yang sedang berjalan, penulis mengusulkan untuk membuat sistem pencarian kata kamus bahasa Indonesia. Dengan sistem ini diharapkan pengguna dapat merasa nyaman dan lebih efisien dalam hal pencarian dengan sistem yang telah diusulkan ini.



Gambar 2. Analisa Sistem Usulan

3.2 Perancangan Algoritma

Perancangan algoritma pada sistem ini memiliki algoritma *Levenshtein Distance*. Flowchart program editor kata ditunjukkan pada gambar berikut:



Gambar 3. Flowchart *Levenshtein Distance*

3.3 Rancangan Algoritma *Levenshtein Distance*

Proses *Levenshtein Distance* dilakukan dengan membuat matriks dari dua kata yang dibandingkan (kata yang salah dengan kata baku). Dari setiap kata yang salah dicari jaraknya dengan seluruh kata baku yang ada di database dan nilai *Levenshtein Distance*. Algoritma proses *Levenshtein Distance* ditunjukkan pada pseudocode berikut :

```

Pseudocode Levenshtein Distance

function difference_word($word_from_user, $word_from_db){
    $matrix_track = [];
    for ($i = 0; $i <= strlen($word_from_user); $i++) {
        $matrix_track[0][$i] = $i;
    }
    for ($j = 0; $j <= strlen($word_from_db); $j++) {
        $matrix_track[$j][0] = $j;
    }
    for ($j = 1; $j <= strlen($word_from_db); $j++) {
        for ($i = 1; $i <= strlen($word_from_user); $i++) {
            $indicator = $word_from_user[$i - 1] == $word_from_db[$j - 1] ? 0 : 1;
            $matrix_track[$j][$i] = min($matrix_track[$j][$i - 1] + 1, $matrix_track[$j - 1][$i] + 1,
            $matrix_track[$j - 1][$i - 1] + $indicator);
        }
    }
}
  
```

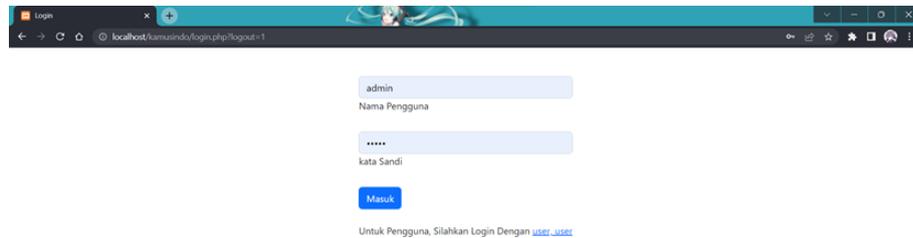
```
}  
}  
  
return $matrix_track[strlen($word_from_db)][strlen($word_from_user)];  
  
}
```

4. IMPLEMENTASI

4.1 Implementasi Antar Muka (*User Interface*)

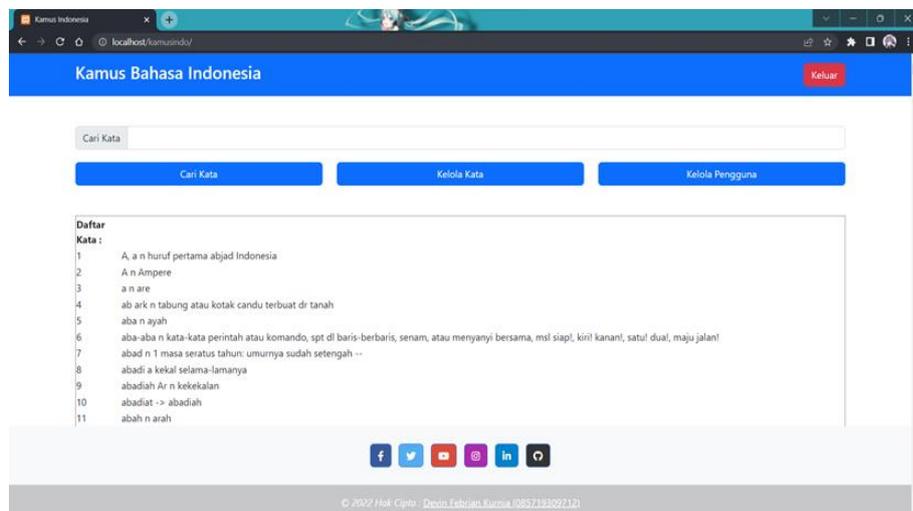
Tujuan implementasi adalah untuk menerapkan perancangan yang telah dilakukan terhadap sistem sehingga user dapat memberi masukan demi berkembangnya sistem yang telah dibangun sebagai simulasi dari sistem pencarian kata kamus besar bahasa Indonesia. Implementasi antarmuka menggambarkan tampilan dari sistem yang dirancang. Berikut ini adalah implementasi antarmuka dari sistem yang dibuat :

4.2 Implementasi Tampilan *Login*



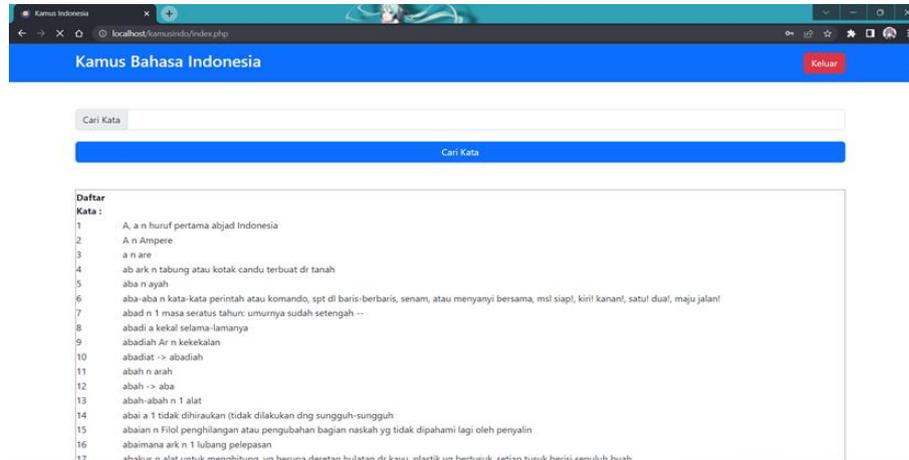
Gambar 4. Implementasi Tampilan *Login*

4.3 Implementasi Tampilan *Index Admin*



Gambar 5. Implementasi Tampilan *Index Admin*

4.4 Implementasi Tampilan *Index User*



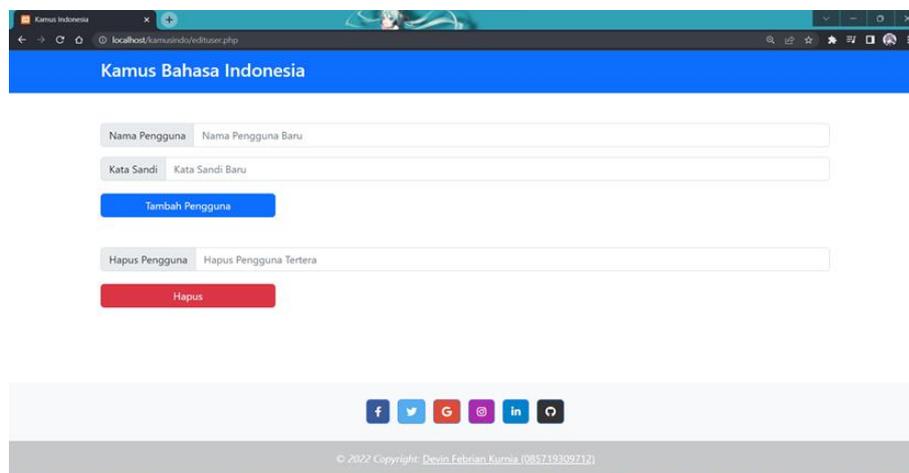
Gambar 6. Implementasi Tampilan *Index User*

4.5 Implementasi Tampilan Kelola Kata



Gambar 7. Implementasi Tampilan Kelola Kata

4.6 Implementasi Tampilan Kelola *User*



Gambar 8. Implementasi Tampilan Kelola *User*

5. KESIMPULAN

Berdasarkan hasil penelitian, analisis, perancangan, implementasi, dan pengujian sistem ini, maka kesimpulan yang didapat adalah sebagai berikut:

1. Penelitian ini menghasilkan sistem pencarian kata pada kamus Bahasa Indonesia menggunakan metode *Approximate String Matching* berbasis *web* dengan bahasa pemrograman *PHP* dan basis data *MySQL*
2. Dalam penggunaannya, sistem ini mampu menghasilkan informasi tentang kata baku yang belum banyak diketahui para pengguna dengan waktu yang cukup efisien.
3. Berdasarkan penelitian yang dilakukan dengan beberapa tahap pengujian pencarian kata menggunakan algoritma *Approximate String Matching*, bahwa penerapan metode pada sistem ini berhasil dilakukan, dapat dilihat dari beberapa penginputan kata dan menghasilkan *output* yang sesuai.

REFERENCES

- Ardi Isbad Amar Gurning, Zanelly , and Arabiatul Adawiyah. (2016). "Penerapan Fuzzy String Matching Pada Aplikasi Pencarian Tugas Akhir Mahasiswa Jurusan Sistem Informasi Berbasis Web," Jurnal Rekayasa dan Manajemen Sistem Informasi.
- Argakusumah, K.W. & Hansun, S. (2014). Implementasi Algoritma Boyer Moore pada Aplikasi Kamus Kedokteran Berbasis Android.
- Fitriana, L. (2021). Implementasi Algoritma Fuzzy String Matching untuk Aplikasi Dispensasi bagi Guru dan Pegawai Berbasis Android di MAN 2 Ponorogo (Doctoral dissertation, Universitas Muhammadiyah Ponorogo).
- Friendly Friendly, (2018). Perbaikan Metode Jaro–Winkler Distance Untuk Approximate String Search Menggunakan Data Terindeks Aplikasi Multi User.” Jurnal Teknovasi : Jurnal Teknik dan Inovasi 4, no. 2.
- Herlawati, P. P. W., & Widodo, P. P. (2012). Menggunakan UML. Bandung: Informatika.
- Ilmy, M. B., Rahmi, N., & Bu'ulölö, R. L. (2016). Penerapan Algoritma Levenshtein Distance untuk Mengoreksi Kesalahan Pengejaan pada Editor Teks. Departemen Teknik Informatika, Institut Teknologi Bandung
- Khairunnisa, K. (2019). Penerapan Fuzzy String Matching Pada Aplikasi Pencarian Judul Skripsi Teknik Informatika Berbasis Android. JURIKOM (Jurnal Riset Komputer), 6(3), 296-300.
- Kuswanto, W. (2020). Implementasi Algoritma Levenshtein Distance Dengan Restful Web Service Pada Kata Bahasa Indonesia Ke Bahasa Jawa Berbasis Web. JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia), 5(2), 78-85.
- Mustaqbal. (2015). Pengujian Aplikasi Menggunakan Black Box Testing Boundary Value Analysis. Bandung: Informatika.
- Raharjo, B., & Heryanto, I. (2012). Modul Pemrograman Web html, php & mysql.
- Rizky, S. (2013). Konsep Dasar Rekayasa perangkat Lunak. Jakarta: PT. Prestasi Pustakaraya.
- Suhartono, E. (2014). Perancangan Dan Pembuatan Program Aplikasi Pengecekan Struktur Bahasa Dalam Bahasa Indonesia. Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Kristen Petra, Surabaya..