

KOREKSI KESALAHAN EJAAN PADA PENCARIAN ARTIKEL PARIWISATA BERBAHASA INDONESIA MENGGUNAKAN *LEVENSHTEIN DISTANCE*

QUERY SPELLING CORRECTION FOR INDONESIAN TOURISM ARTICLE USING LEVENSHTEIN DISTANCE

¹Ika Oktavia Suzanti*, ²Yudha Dwi Putra Negara², ³Aisyatur Radiah³, ⁴Andharini Dwi Cahyani⁴, ⁵Husni⁵,

^{1,2,3,4,5} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo Madura
Bangkalan, Madura, Indonesia

*e-mail: iosuzanti@trunojoyo.ac.id

Abstrak

Perkembangan teknologi informasi di era digital saat ini, membawa perubahan dan akses informasi yang cepat dan akurat dalam segala aspek kehidupan manusia, termasuk berita pariwisata sebagai salah satu contohnya. Dalam memutuskan untuk berlibur ke suatu tempat, wisatawan membutuhkan referensi tempat pariwisata salah satunya melalui artikel online. Hal tersebut memudahkan wisatawan untuk mencari referensi melalui portal berita pariwisata dengan menggunakan mesin pencari. Spelling correction adalah cara paling umum untuk mengidentifikasi dan memberikan saran atau ide pada kata-kata yang salah ejaannya dalam sebuah teks. Spelling correction dapat digunakan untuk memperbaiki kesalahan ejaan dari query Bahasa Indonesia yang dimasukkan terhadap artikel pariwisata. Metode Levenshtein Distance merupakan salah satu metode yang dapat digunakan untuk memperbaiki kesalahan ejaan kata yang disebabkan oleh kesalahan pengetikan sehingga hasil yang diberikan tidak sesuai dengan query (kata kunci) yang ada. Metode ini bekerja dengan cara membuat matriks untuk mengukur jumlah selisih antara dua string. Hasil dari penelitian ini menunjukkan bahwa penggunaan metode Levenshtein Distance dalam memperbaiki ejaan kata sangat baik dengan nilai presisi 1,0 dan recall 0,89.

Kata kunci: Bahasa Indonesia, Koreksi Ejaan, Levenshtein Distance.

Abstract

The information technology development in current digital era, brings changes and fast and accurate access to information in all aspects of human life, including tourism news as one example. In deciding to go on vacation to a place, tourists need references to tourist attractions, one of which is through online articles. This makes it easier for tourists to find references through tourism news portals using search engines. Spelling correction is the most common way to identify and provide suggestions or ideas for misspelled words in a text. Spelling correction can be used to correct spelling errors from Indonesian queries entered into tourism articles. The Levenshtein Distance method is one method that can be used to correct spelling errors caused by typing errors so that the results given do not match the existing query (keywords). This method works by creating a matrix to measure the amount of difference between two string s. The results of this study indicate that the use of the Levenshtein Distance method in correcting word spelling is very good with a precision value of 1.0 and a recall of 0.89.

Keywords: Indonesian language, Spelling Correction, Levenshtein Distance.

1 PENDAHULUAN

Perkembangan teknologi merupakan hal umum yang dapat ditemukan dalam berbagai aspek kebutuhan manusia, yaitu sebagai alat komunikasi, visualisasi, pendidikan, bisnis, dan lain-lainnya [1]. Berita adalah cerita atau uraian tentang suatu kejadian atau peristiwa yang bersifat factual, penting, dan diminati sebagian besar pembaca, serta berkaitan dengan berbagai kepentingan. Berita daring terbagi menjadi beberapa bagian, salah satunya adalah artikel yang diterbitkan untuk menyampaikan informasi pariwisata [2]. Pariwisata merupakan industri yang penting karena memuat informasi tentang pariwisata seperti destinasi, *event*, dan hotel. Adanya aktivitas tersebut menjadi salah satu permintaan yang umum dalam pencarian daring. Salah satu cara wisatawan dapat mencari referensi adalah melalui portal berita pariwisata dengan menggunakan mesin pencari. Dalam hal ini, hanya dengan menginputkan *query* pada mesin pencari, sistem kemudian memberikan umpan balik berdasarkan *query* yang diberikan [3]. Namun seringkali hasil yang didapatkan wisatawan tidak relevan dengan apa yang diharapkan, beberapa penyebabnya adalah kesalahan dalam memberikan *query* yang tidak mencantumkan ejaan baku atau terdapat kesalahan ketik.

Pembetulan ejaan kata, apabila ditemukan kata yang tidak tepat, maka harus dilakukan pencarian kata yang mungkin padanannya [4]. Adanya kesalahan ejaan dapat menghambat pencarian informasi yang akurat dan relevan, terutama ketika mesin pencari tidak dapat memahami kata yang salah eja. Oleh karena itu, terdapat satu topik untuk mengatasi masalah kesalahan ejaan dalam pencarian informasi menggunakan *query*, yaitu pembetulan ejaan. Pembetulan kesalahan ejaan merupakan fitur pembetulan otomatis yang digunakan untuk melakukan proses pendeteksian dalam kesalahan ejaan kata dan juga memberikan saran kata dalam kesalahan ejaan pada teks [5]. Pembetulan ejaan terbagi menjadi dua jenis, yaitu: pembetulan kesalahan non-kata dan pembetulan kesalahan kata aktual. Pembetulan kesalahan non-kata berfokus pada penanganan kata yang salah eja yang disebabkan oleh kesalahan tipografi [6], [7].

Berdasarkan kesalahan ejaan pada artikel berita bahasa Indonesia di atas, maka diperlukan suatu metode pembetulan ejaan [8]. Untuk permasalahan tersebut, pembetulan ejaan dilakukan dengan menggunakan *Levenshtein Distance* yang dapat membantu pembetulan kesalahan ejaan dengan cara mencari nilai jarak (*Edit Distance*) kata salah yang terdeteksi dengan semua kata baku yang ada di kamus [9]. *Levenshtein Distance* merupakan salah satu cara pencarian solusi dalam bidang komputasi bahasa alami, untuk dapat menentukan penyelesaian salah satunya adalah menentukan matriks untuk mengukur besarnya selisih dua *string* [4]. Untuk mempermudah proses pembetulan ejaan, telah ada penelitian terkait yang berhasil diterapkan, terdapat beberapa metode dalam pembetulan ejaan antara lain, Beberapa penelitian yang membahas tentang spelling check lebih banyak membahas tentang kesalahan kata yang mana disebabkan oleh kesalahan tipografi. Penelitian yang dilakukan oleh Tria Aprilianto, Achmad Badawi, pada tahun 2017, dengan menggunakan metode pendekatan kamus berbasis *Levenshtein Distance*, pada sistem koreksi kata dan pengenalan struktur kalimat Bahasa Indonesia, dapat disimpulkan bahwa aplikasi tidak dapat memproses teks lebih dari 1000 kata atau 6 halaman, dengan pengujian kesalahan ejaan kata dilakukan dengan memasukkan kata yang salah. Dan hasilnya menunjukkan bahwa saran kata yang diprioritaskan adalah kata yang benar, dan pengujian kesalahan kata memiliki akurasi sebesar 86%. Dan pengujian struktur bahasa Indonesia dilakukan dengan memberikan kalimat dan output sistem kepada penguji sebanyak 30 kalimat dengan nilai akurasi sebesar 76,66% [9]. Dan penelitian yang berjudul Autocorrect pada Modul Pencarian e-Dictionary Obat Menggunakan Algoritma *Levenshtein Distance*, yang ditulis oleh Halimah Tus Sadiyah, dkk. Hasil akurasi pengujian penerapan algoritma pada dinas farmasi menghasilkan akurasi, recall dan presisi sebesar 90% [10].

Beberapa alasan di atas, maka dibuatlah sebuah sistem untuk melakukan *Spelling correction* atau perbaikan kesalahan ejaan terhadap *query* dalam pencarian artikel atau berita pariwisata berbahasa Indonesia. Untuk mengatasi permasalahan yang ada pada penelitian *Spelling correction* sebelumnya, maka diusulkan penggunaan Algoritma *Levenshtein Distance*.

2 TINJAUAN PUSTAKA

Spelling correction adalah cara paling umum untuk mengidentifikasi dan memberikan saran atau ide pada kata-kata yang salah ejaannya dalam sebuah teks. Koreksi ejaan adalah sebuah sistem yang digunakan untuk dapat melakukan pengoreksian kesalahan dalam ejaan [11]. Kesalahan yang sering terjadi yaitu adanya karakter yang tidak sesuai, misalnya terdapat karakter yang berubah, atau kurangnya karakter dan terdapat penggantian karakter [12]. Misalnya dalam sebuah kata yang salah yaitu “Keuar”, kemungkinan kata yang benar adalah “Keluar”, penyebab kesalahan katanya adalah kurang huruf “l”.

Algoritma ini merupakan sebuah algoritma yang digunakan untuk koreksi ejaan yaitu algoritma *Levenshtein Distance* yang ditemukan oleh seorang ilmuwan dari Rusia pada tahun 1965 yang bernama Vladimir Levenshtein, dimana algoritma ini merupakan algoritma yang menggunakan matriks untuk mengukur perbedaan antara dua *string*. Pada algoritma Levenshtein Distance, dua buah *string* merupakan jumlah dari minimum operasi yang dibutuhkan untuk mengubah sebuah *string* (*source string*) menjadi *string* yang lain (*string target*). Dalam algoritma Levenshtein Distance ini merupakan suatu operasi yang melibatkan *insertion* (penyisipan), *deletion* (penghapusan), *substitution* (penggantian) dari suatu karakter tunggal [13][13]. Berikut ini merupakan penjelasan dari operasi algoritma *Levenshtein Distance* sebagai berikut:

1. Operasi Penyisipan Karakter (*Insertion*)

Insertion merupakan sebuah operasi penyisipan karakter baru kedalam *string* menggunakan persamaan 2. Dalam proses penyisipan karakter, dapat dilakukan di awal kalimat, maupun akhir kalimat.

2. Operasi Penghapusan Karakter (*Deletion*)

Deletion adalah operasi penghapusan karakter yang berlebihan dalam suatu karakter seperti pada persamaan 1. Contohnya kata “matematikan” nantinya dilakukan penghapusan sebuah kata yaitu hasilnya menjadi “matematika”.

3. Operasi perubahan Karakter (*Substitution*)

Substitution yaitu operasi yang melakukan pergantian dalam suatu karakter dengan karakter lainnya yang bernilai benar seperti pada persamaan 3. Misalnya sebuah kata “yang” diganti menjadi kata “yng” dalam operasi perubahan sebuah karakter yang dimana pada operasi ini nantinya huruf “m” diganti menjadi “n”.

Perhitungan *Levenshtein Distance* ini didapat pada sebuah matriks yang digunakan dalam menghitung dari jumlah perbedaan yang terdapat dalam dua *string*. Pada perhitungan jarak antara dua *string* yang ditentukan dari jumlah minimum operasi pada perubahan untuk membuat sebuah *string A* menjadi *string B* [13]. Fungsi Algoritma *Levenshtein Distance* yaitu matriks 2 dimensi yang digunakan untuk perhitungan nilai jarak dalam *Levenshtein Distance*. Dengan isi di dalam nilainya yaitu matriks yang mempunyai jumlah operasi penghapusan, penyisipan, dan penggantian yang diperlukan dalam mengubah sebuah *string target*. Berikut ini rumus algoritma *Levenshtein Distance* [14].

$$lev_{a,b}(i, j) = \min lev_{a,b}(i - 1, j) + 1 \text{ (Penghapusan)} \quad (1)$$

$$lev_{a,b}(i, j) = \min lev_{a,b}(i, j - 1) + 1 \text{ (Penyisipan)} \quad (2)$$

$$lev_{a,b}(i, j) = \min lev_{a,b}(i - 1, j - 1) + 1, ai \neq bj \text{ (Substitusi)} \quad (3)$$

$$lev_{a,b}(i, j) = \min lev_{a,b}(i - 1, j - 1), ai = bj \text{ (Tidak ada perubahan)} \quad (4)$$

Keterangan:

a = *string* Sumber

b = *String* Target

i = *index* baris *string* sumber

j = *index* baris *string* target

$(ai \neq bj)$ = perlu menambahkan 1 (+1)

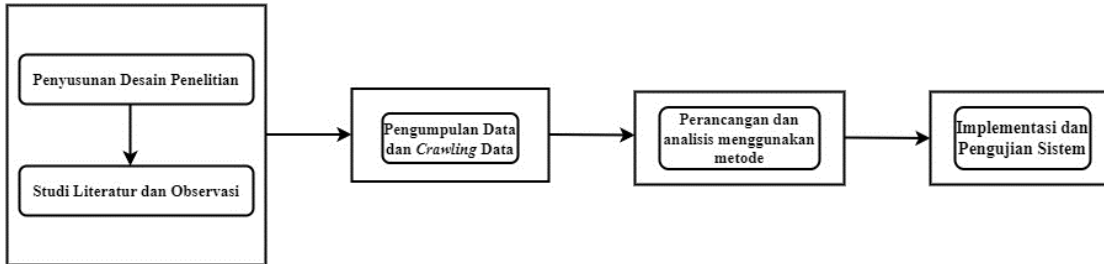
Kondisi $ai \neq bj$

$(ai = bj)$ = tidak perlu menambahkan 1 (+1)

3 METODE PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian pengembangan sistem koreksi kesalahan ejaan pada *query* menggunakan *Levenshtein Distance* dapat dilihat pada Gambar 1 yaitu tahapan penelitian.



Gambar 1. Tahapan Penelitian

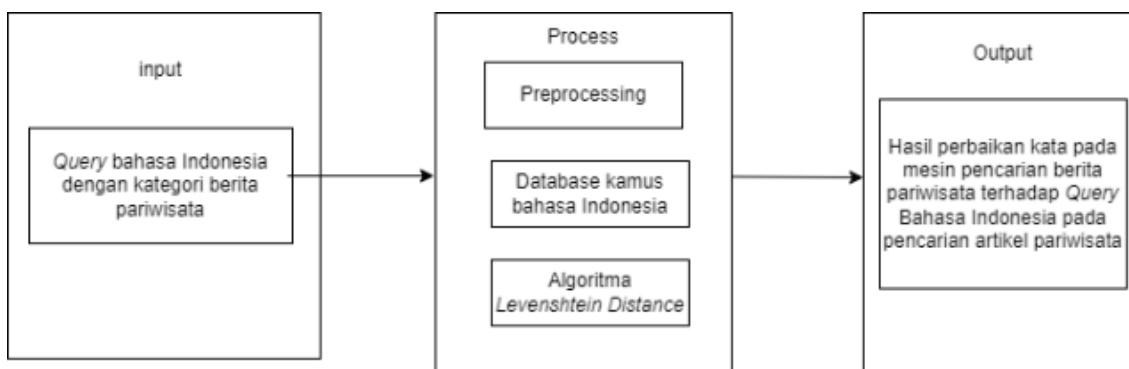
Tahap penelitian yang dilakukan dalam merancang sistem penelitian ini yaitu dengan melakukan tahap studi pustaka dan mengumpulkan data serta informasi yang membantu dalam penyelesaian penelitian. Data dan informasi yang dikumpulkan meliputi materi tentang metode yang digunakan seperti jurnal tentang *Spelling correction* dan *Levenshtein Distance*. Kemudian, pengumpulan data dengan cara crawling data artikel berita pariwisata, dan juga kamus bahasa Indonesia. Data yang diperoleh diolah ke tahap preprocessing. Selanjutnya, menentukan, menganalisis dan merancang sistem *Spelling correction query* bahasa Indonesia dalam pencarian artikel berita pariwisata. Pada perancangan dan analisis menggunakan search engine untuk *Spelling correction* yaitu user menginputkan *query* bahasa Indonesia, dimana *Spelling correction* tersebut menggunakan metode *Levenshtein Distance* untuk mengkoreksi ejaan *query* inputan dan sistem menampilkan informasi berita pariwisata sesuai dengan *query* bahasa Indonesia yang diinputkan.

3.2 Data

Penelitian ini, data dikumpulkan dari laman Detik.com. Data yang digunakan adalah data artikel berita pariwisata yang diperoleh dari hasil crawling pada situs web Detik.com atau laman dengan kategori berita pariwisata yang berjumlah 332 dokumen dengan jumlah kata 133.403. Dalam hal ini, dataset yang digunakan adalah dataset bahasa Indonesia dengan kategori artikel berita pariwisata.

3.3 Desain Penelitian

Tahap desain penelitian, terbagi menjadi Input, Proses, dan Output. Berikut penjelasan singkat mengenai IPO (Input, Proses, dan Output) pada Gambar 2.

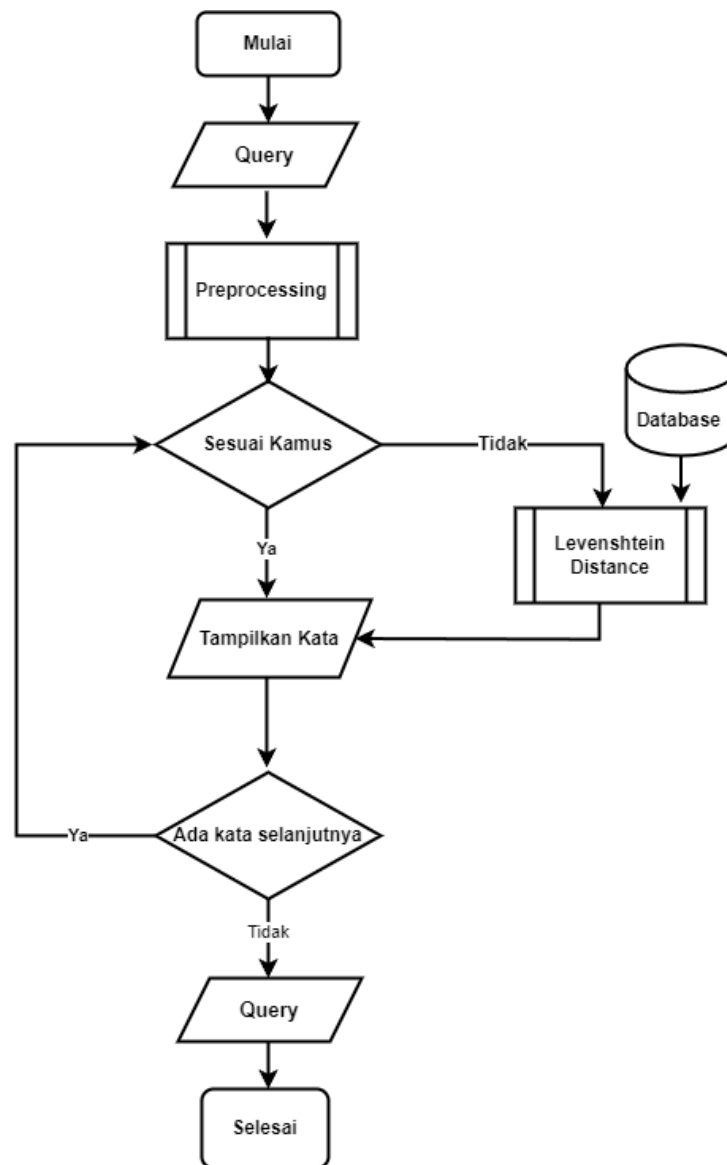


Gambar 2. Diagram IPO

3.3.1 Diagram Alir Sistem

Diagram alir sistem merupakan representasi dari suatu aliran dalam sistem yang dirancang untuk membantu pemahaman konsep algoritma dalam suatu sistem. Diagram alir ini merepresentasikan algoritma dalam proses perbaikan kesalahan ejaan pada *query* pencarian

artikel pariwisata berbahasa Indonesia dengan menggunakan algoritma *Levenshtein Distance*. Pada bagian ini merupakan diagram alir dalam menentukan perbaikan kesalahan ejaan pada *query* artikel pariwisata berbahasa Indonesia, dengan cara mengecek pada kamus bahasa Indonesia yang digunakan.



Gambar 3. Cara kerja *Levenshtein Distance*

Tahapan proses yang terdapat pada Gambar 3 yaitu: diawali dengan *crawling* data yang menampilkan data *output* dari hasil *preprocessing*. Langkah pertama diawali dengan user memasukkan kata *input* berupa *query* Bahasa Indonesia, dalam hal ini sebelum melakukan *word repair* dilakukan proses *preprocessing* dimana terdapat 3 (tiga) tahap yaitu *Case Folding*, *cleaning*, dan *tokenizing*. Selanjutnya jika kata *input* berupa *query* Bahasa Indonesia sesuai dengan kamus maka ditampilkan kata yang terdapat pada kamus. Jika kata *input* tidak ada pada kamus kata maka dilakukan pengecekan pada kamus kata yang memiliki panjang karakter yang sama dengan kata input. Kata yang tidak ada pada kamus dianggap sebagai kata yang salah dan dilakukan proses perhitungan *Levenshtein Distance*.

3.3.2 Pre-processing Data

Tahap ini dilakukan *preprocessing*, agar kumpulan data siap digunakan. Data mentah diolah pada tahap *preprocessing*. Proses *preprocessing* teks ini digunakan untuk mengurangi noise dan merapikan data, menyamakan bentuk kata, serta mengurangi volume data. Sehingga nantinya data yang sudah diolah siap untuk diolah, sedangkan data yang didapatkan dari hasil *crawling* dihapus.

Seperti yang terlihat pada tabel di bawah ini, berikut ini adalah proses dari tahap preprocessing data.

1. *Case Folding*

Case Folding adalah proses mengubah huruf kapital menjadi huruf kecil, di mana setiap karakter az dalam kata dan kalimat diubah menjadi huruf kecil [15]. Berikut ini adalah contoh hasil *Case Folding* pada tabel 1.

Tabel 1. Hasil *Case Folding*

Kata Sebelum <i>Case Folding</i>	Kata setelah <i>Case Folding</i>
Karnaval Budaya Klaten 2023 Memanfaatkan Berbagai Potensi Seni Budaya	karnaval budaya klaten 2023 menggunakan berbagai potensi seni budaya

2. *Cleaning*

Cleaning merupakan proses pembersihan karakter-karakter tertentu yang tidak diperlukan seperti simbol dan angka pada kata dan kalimat [15]. Berikut ini adalah contoh hasil *cleaning* pada Tabel 2.

Tabel 2. Hasil *Cleaning*

Kata Sebelum Membersihkan	Kata Setelah Pembersihan
karnaval budaya klaten 2023 menggunakan berbagai potensi seni budaya	karnaval budaya klaten menggunakan berbagai potensi seni budaya

3. *Tokenisasi*

Tokenisasi adalah proses memecah kalimat menjadi unit terkecil (kata/token) [15]. Berikut ini adalah contoh hasil Tokenisasi pada tabel 3 di bawah ini.

Tabel 3. Hasil Tokenisasi

Kata Sebelum Tokenisasi	Kata Setelah Tokenisasi
karnaval budaya klaten menggunakan berbagai potensi seni budaya	['karnaval','budaya', 'klaten', 'usung', 'berbagai', 'potensi', 'seni', 'budaya']

4. Implementasi Sistem

Tahap implementasi sistem dilakukan dengan membangun sistem yang sesuai dengan perancangan dan hasil analisis kebutuhan sistem yang telah ditentukan sebelumnya. Pada implementasi sistem ini dibahas mengenai sistem seperti apa yang nantinya dibangun dalam menyelesaikan permasalahan koreksi ejaan kata menggunakan *Levenshtein Distance* pada sistem temu kembali informasi bahasa Indonesia dalam pencarian artikel berita pariwisata.

4 HASIL DAN PEMBAHASAN

4.1. Kumpulan data

Dataset yang digunakan adalah dataset bahasa Indonesia yang diperoleh dari hasil crawling pada situs web atau laman Detik.com dengan kategori berita pariwisata sebanyak 332 dokumen dengan 133.403 kata yang disimpan dalam file csv. Tabel 4. merupakan dataset yang digunakan.

Tabel 4. 1Dataset Bahasa Indonesia

TIDAK	Kata
1	aba
2	abad
3	abadi
4	abah
...	...
...	...
23903	ke dalam

TIDAK	Kata
23904	senam
23905	zuna
23906	zuroh
23907	zusnali

4.2. Data Uji

Data uji ini dilakukan dengan melakukan *query* salah dan *query* benar yang nantinya dilakukan perbaikan ejaan pada *query* bahasa Indonesia dalam pencarian artikel berita pariwisata. *Query* ini berkaitan dengan pariwisata seperti ditunjukkan pada table 5. Data uji yang digunakan dalam penelitian ini berjumlah 10 data *query* salah tentang pariwisata. Kesalahan kata pada data uji berjumlah 18 kata dengan rincian penambahan sebanyak 7 kata, penghapusan sebanyak 3 kata, dan penggantian sebanyak 8 kata.

Tabel 5. *2Query Testing*

No.	Kueri yang Salah	Kueri Benar
1.	Destinasi wista indinesia terbaru	Destinasi wisata Indonesia terbaru
2.	Promo Linuran Indonesia	Promo Liburan Indonesia
3.	Acara pariwisata indonesia	Event Pariwisata Indonesia
4.	Kuliner halal bali	Kuliner halal bali
5.	Objek wista tersembunyi di Indonesia	Objek Wisata Tersembunyi di Indonesia
6.	Pandian perjalanan ke bali	Panduan perjalanan ke bali
7.	Wisata murah di Indonesia	Wisata murah di Indonesia
8.	Harga tiket wisata malam	Harga tiket wisata malang
9.	Harga swa bis pariwisata bekasi	Harga sewa bus pariwisata bekasi
10.	Tujuan wisata tebraru	Destinasi Wisata Terbaru

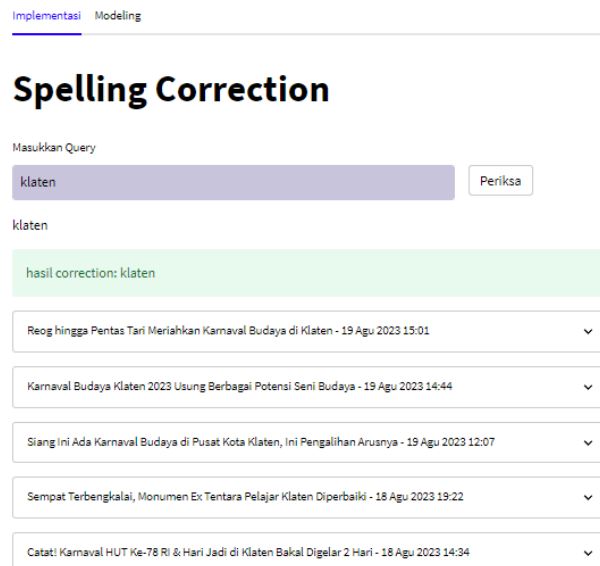
4.3. Hasil Implementasi Sistem

Implementasi merupakan tahap ketika suatu rancangan telah rampung, kemudian diimplementasikan lebih lanjut ke dalam bahasa pemrograman yang digunakan. Implementasi sistem ini merupakan tahap penerapan sistem agar nantinya sistem tersebut dapat dioperasikan. Implementasi sistem berfungsi untuk mengonfirmasi modul-modul rancangan, sehingga nantinya pengguna dapat memberikan masukan kepada pengembang sistem.

4.4. Hasil Evaluasi

4.4.1. Pengujian

Pengujian sistem dilakukan untuk mengukur tingkat keberhasilan sistem dalam mengoreksi kata yang dilakukan dalam 2 tahap, yaitu tahap pengujian kata berdasarkan data latih, dan tahap pengujian kata terhadap input pengguna.



Gambar 4. Inputan *query* pada sistem

a) Hasil Pengujian Data Pelatihan

Pengujian koreksi kata atau *query* dilakukan dengan menggunakan data latih yang bersumber dari data latih penambahan karakter, data latih penghapusan karakter, dan data latih penggantian karakter.

b) Hasil Pengujian Input Pengguna

Pengujian koreksi kata berdasarkan masukan pengguna dilakukan sesuai dengan kebutuhan pengguna seperti pada gambar 4. Pengguna bebas memasukkan pertanyaan, dan sistem hanya menampilkan saran koreksi.

c) Hasil Pengujian Presisi dan Recall

Hasil pengujian precision dan recall dari 10 data uji *query* yang digunakan sebagai data uji diperoleh hasil precision dan recall yaitu untuk precision 1.0 dan recall 0.89 seperti ditunjukkan pada gambar 5.

Hasil Koreksi kata pada Data test

	Query Pengujian	Hasil Koreksi Kata
0	Destinasi wista indonesia terbaru,?"	destinasi wisata indonesia terbaru
1	Promo linuran Indonsia	promo liburan indonesia
2	Event Pariwisata Indonesua	event pariwisata indonesia
3	Kulinr halal bali	kuliner halal bali
4	Wista Tersembunyi di Indonesia	wisata tersembunyi di indonesia
5	Balp Ssepeda sambill bewisata	bala sepeda sambil berwisata
6	wista murah di indonesia	wisata murah di indonesia
7	harga tiket wisata malang	harga tiket wisata malang
8	bus wista terbkar hehatt di jaln Solo	bus wisata terbakar hebat di jln solo
9	Klaten menyelnggarakan karnala budya	klaten menyelenggarakan karbala budaya

Hasil Precision dan Recall

Recall: 0.8945238095238095

Precision: 1.0

Gambar 5. Hasil koreksi *query***5 KESIMPULAN**

Berdasarkan hasil pengujian dan pembahasan penelitian yang telah diuraikan sebelumnya, penelitian ini telah memberikan hasil sesuai dengan hipotesisnya yaitu

1. Analisis koreksi ejaan pada kueri artikel berita pariwisata berbahasa Indonesia menggunakan metode *Levenshtein Distance* dapat diterapkan dengan baik untuk pencarian data input. Aplikasi ini bukan untuk menentukan hasil akhir, melainkan hanya proses sistem dalam mengoreksi kesalahan ejaan dan hanya memberikan rekomendasi.
2. Hasil presisi dan recall dari 10 *query* data uji yang dibuat untuk dijadikan data latih dengan melakukan perhitungan presisi dan recall dengan mengunggah file csv ke sistem diperoleh presisi 1,0 dan recall 0,89.
3. Penggunaan metode *Levenshtein Distance* dari hasil uji presisi dan akurasi recall serta penerapan *query* pencarian artikel berita pariwisata berbahasa Indonesia menunjukkan bahwa metode *Levenshtein Distance* dapat diterapkan pada sistem ejaan kata bahasa Indonesia, dan hasil perhitungan seluruh data uji juga baik.

UCAPAN TERIMA KASIH

Terima kasih kepada LPPM Universitas Trunonojoyo Madura

DAFTAR PUSTAKA

- [1] H. Hartono, A. Hajjah, and Y. N. Marlim, "Penerapan Metode Naïve Bayes Classifier Untuk Klasifikasi Judul Berita," *Jurnal Simantec*, vol. 12, no. 1, pp. 37–46, 2023.
- [2] A. Y. R. Fitriani and L. E. Rahmawati, "Analisis kesalahan penggunaan tanda baca dan huruf miring dalam teks berita online detiknews dan tribunnews," *Bahastra*. DOI: <https://doi.org/http://dx.doi.org/10.26555/bahastra.v40i1>, vol. 14695, 2020.
- [3] N. Ahmad, A. A. Prasetyo, and A. Masruri, "Penerapan Information Retrieval Pada Search Engine," *Knowledge: Jurnal Inovasi Hasil Penelitian dan Pengembangan*, vol. 1, no. 1, pp. 15–23, 2021.
- [4] M. O. Braddley, M. Fachrurrozi, and N. Yusliani, "Pengoreksian Ejaan Kata Berbahasa Indonesia Menggunakan Algoritma Levensthein Distance," in *Annual Research Seminar (ARS)*, 2017, pp. 167–172.
- [5] C. Prianto, D. Markuci, and S. F. Pane, "Implementasi Spelling Corrector Untuk Mengatasi Typographical Error Pada Fitur Pencarian Aplikasi Kamus Istilah Informatika," *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 17, no. 1, pp. 20–30, 2023.
- [6] M. S. Simanjuntak, H. Sujaini, and N. Safriadi, "Spelling corrector bahasa indonesia dengan kombinasi metode peter norvig dan n-gram," *Jurnal Edukasi dan Penelitian Informatika*, vol. 4, no. 1, p. 17, 2018.
- [7] A. I. Fahma, I. Cholissodin, and R. S. Perdana, "Identifikasi kesalahan penulisan kata (typographical error) pada dokumen berbahasa Indonesia menggunakan metode N-gram dan Levenshtein Distance," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, pp. 53–62, 2018.
- [8] Y. Yanfi, F. L. Gaol, B. Soewito, and H. Warnars, "Spell Checker for the Indonesian Language: Extensive Review," *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, no. 5, pp. 1–7, 2022.
- [9] T. Aprilianto and A. Badawi, "Sistem Koreksi Kata Dan Pengenalan Struktur Kalimat Berbahasa Indonesia Dengan Pendekatan Kamus Berbasis Levenshtein Distance," *SPIRIT*, vol. 9, no. 1, 2018.

- [10] H. T. Sadiyah, M. S. N. Ishlah, N. N. Rokhmah, and others, “Autocorrect on Drugs e-Dictionary Search Module Using Levenshtein Distance Algorithm,” *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 4, no. 1, pp. 64–69, 2020.
- [11] Y. Apriani, S. M. Solikhah, and A. B. Prapsita, “Kesalahan Ejaan Bahasa Indonesia Pada Berita Koran Kedaulatan Rakyat Edisi Mei 2021,” *Jurnal Kajian Bahasa dan Sastra Indonesia*, vol. 10, no. 3, pp. 230–236, 2021.
- [12] Y. P. Sari, G. A. Pradnyana, and I. M. A. Wirawan, “Pengembangan Aplikasi Kamus Bahasa Bima-Bahasa Indonesia Menggunakan Algoritma Levenshtein Distance Sebagai Spell Checker Berbasis Android,” *KARMAPATI (Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika)*, vol. 8, no. 2, pp. 86–95, 2019.
- [13] D. Rosmala and Z. M. Risyad, “Algoritma Levenshtein Distance dalam Aplikasi Pencarian isu di Kota Bandung pada Twitter,” *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, vol. 2, no. 2, pp. 1–12, 2017.
- [14] D. Hládek, J. Staš, and M. Pleva, “Survey of automatic spelling correction,” *Electronics (Basel)*, vol. 9, no. 10, p. 1670, 2020.
- [15] N. Hamidah, N. Yusliani, and D. Rodiah, “Spelling Checker using Algorithm Damerau Levenshtein Distance and Cosine Similarity,” *Sriwijaya Journal of Informatics and Applications*, vol. 1, no. 1, 2020.