

**KOREKSI EJAAN KATA BAHASA MELAYU ROKAN HULU  
MENGUNAKAN METODE *LEVENSHTEIN DISTANCE***

**SKRIPSI**

**OLEH :**

**RAMA WARDANA**  
**NIM : 1737018**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PASIR PENGARAIAN**

**2022**

**PERSETUJUAN PEMBIMBING**  
**KOREKSI EJAAN KATA BAHASA MELAYU ROKAN HULU**  
**MENGGUNAKAN METODE *LEVENSHTEIN DISTANCE***

---

Disetujui oleh :

Pembimbing I



**Budi Yanto, S.T., M.Kom**  
NIDN. 1029058301

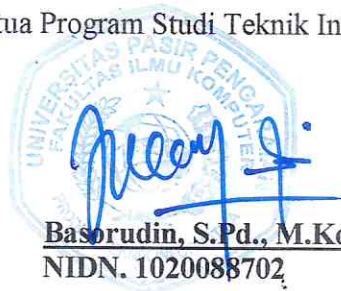
Pembimbing II



**Satria Riki Mustafa, S.Pd., MSi**  
NIDN.1001039301

Deketahui Oleh :

Ketua Program Studi Teknik Informatika








**Basorudin, S.Pd., M.Kom**  
NIDN. 1020088702

## PERSETUJUAN PENGUJI


Tugas Akhir ini telah diuji oleh  
Tim Penguji Ujian Sarjana Komputer  
Program Studi Teknik Informatika  
Fakultasi Ilmu Komputer  
Universitas Pasir Pengaraian  
Pada Tanggal 25 Juli 2022

---

Tim Penguji :

- |   |            |   |
|---|------------|---|
| 1. <u>Budi Yanto, ST., M.Kom</u><br>NIDN.102905831            | Ketua      | (  )   |
| 2. <u>Satria Riki Mustafa, S.Pd., M.Si</u><br>NIDN.1001039301 | Sekretaris | (  )  |
| 3. <u>Erni Rouza, S.T., M.Kom</u><br>NIDN.1009058707          | Anggota    | (  ) |
| 4. <u>Luth Fimawahib, M.Kom</u><br>NIDN.1013068901            | Anggota    | (  ) |
| 5. <u>Basorudin, S.Pd., M.Kom</u><br>NIDN.1020088702          | Anggota    | (  ) |

Mengetahui  
Dekan Fakultas Ilmu Komputer  
Universitas Pasir Pengaraian

  
Hendri Maradona, M.Kom  
NIDN. 1002038702

## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa Tugas Akhir yang berjudul “Kooreksi Ejaan Kata Bahasa Melayu Rokan Hulu Menggunakan Metode *Levenshtein Distance*”, benar hasil penelitian saya dengan arahan dosen pembimbing dan belum pernah diajukan dalam bentuk apapun untuk mendapatkan gelar kesarjanaan. Dalam Tugas Akhir ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasi orang lain, kecuali secara tertulis dengan jelas dicantumkan dalam naskah dengan menyebutkan referensi yang dicantumkan dalam daftar pustaka. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena Tugas Akhir ini, serta lainnya sesuai norma yang berlaku di perguruan tinggi.

Pasir Pengaraian, 25 Juli 2022  
Yang Membuat Pernyataan



**RAMA WARDANA**  
Nim : 1737018

## KATA PENGANTAR

*Assalamu'alaikum wa rahmatullahi wa barakatuh*

*Alhamdulillah Rabbil Alamin*, Puji syukur kehadiran Allah SWT yang senantiasa melimpahkan rahmat dan karunia-Nya, Sehingga penulis mampu menyelesaikan Skripsi ini dengan baik. Shalawat serta salam terucap buat junjungan kita Rasulullah Muhammad SAW karna jasa beliau yang telah membawa manusia dari zaman kebodohan ke zaman yang penuh dengan ilmu pengetahuan seperti sekarang ini.

Skripsi ini berjudul “**Koreksi Ejaan Kata Bahasa Melayu Rokan Hulu Menggunakan Metode *Levenshtein Distance***” sebagai satu syarat untuk mendapatkan kelulusan pada jurusan Teknik Informatika Universitas Pasir Pengaraian. Banyak sekali pihak yang telah membantu penulis dalam penyusunan Skripsi ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada penulis. Semua itu tentu terlalu banyak bagi penulis untuk membalasnya, namun pada kesempatan ini penulis hanya dapat mengucapkan terimakasih kepada :

1. Allah SWT ,yang dengan rahmat-Nya memberikan semua yang terbaik dan dengan hidayah-Nya memberi petunjuk sehingga dalam penyusunan Skripsi ini berjalan dengan lancar.
2. Rasulullah SAW, yang telah membawa petunjuk bagi manusia agar menjadi manusia paling mulia derajatnya di sisi Allah SWT.
3. Kepada Ayah, Ibu, Adik-adik tercinta, atas kerja keras dan selalu memberikan doa restu yang tidak ternilai harganya yang banyak memberikan semangat, motivasi, dan bimbingan yang terbaik dan limpahan kasih sayang yang tiada henti.
4. Bapak Dr. Hardianto, M.Pd selaku Rektor Universitas Pasir Pengaraian.
5. Bapak Hendri Maradona, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Pasir Pengaraian.

6. Bapak Basorudin, S.Pd., M.Kom selaku Ka. Prodi, selaku Ketua Prodi Teknik Informatika Fakultas Ilmu Komputer Universitas Pasir Pengaraian, sekaligus selaku penguji.
7. Bapak Budi Yanto, S.T., M.Kom selaku pembimbing 1 yang telah memberikan bimbingan, arahan, dan saran yang berharga dalam menyusun Skripsi ini.
8. Bapak Satria Riki Mustafa, S.Pd., M.Si selaku pembimbing 2 yang telah memberikan bimbingan, arahan, dan saran yang berharga dalam menyusun Skripsi ini.
9. Bapak Imam Rangga Bakti, M.Kom sebagai koordinator Skripsi yang telah banyak membantu dalam menyusun jadwal dan koordinasi dengan para pembimbing dan sesuatu hal yang memperlancar jalannya Skripsi ini.
10. Seluruh staf dan pegawai Tata Usaha Fakultas Ilmu Komputer Universitas Pasir Pengaraian yang telah memberikan bantuan dan kelancaran administratif.

Penulis menyadari bahwa dalam penulisan Skripsi ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat penulis harapkan untuk kesempurnaan Skripsi ini. Akhirnya penulis berharap semoga Skripsi ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya. Amin.

Wassalamu'alaikum wa rahmatullahi wa barakatuh

Pasir Pengaraian, 25 Juli 2022

Mahasiswa

**RAMA WARDANA**  
**NIM : 1737018**

## ABSTRAK

Bahasa Melayu Rokan Hulu merupakan bahasa yang memiliki campuran bahasa, ada tiga campuran bahasa yang menjadi dasar bahasa Melayu Hokan Hulu yaitu antara bahasa Mandailing, Melayu, dan Minangkabau. Bahasa Melayu Rokan Hulu digunakan sehari-hari oleh masyarakat Rokan Hulu untuk saling berkomunikasi baik secara lisan mau pun tulisan seperti di media sosial, akan tetapi banyak pengguna di Rokan Hulu memposting dan berkomentar menggunakan ejaan kata bahasa melayu dengan ejaan kata yang kurang tepat, seperti penyingkataan kata, kesalahan penulisan disebut juga dengan *typo*. Hal ini dapat menimbulkan kesalahan ejaan kata bahasa dalam arti ejaan bahasa baku Melayu Rokan Hulu. Dalam peneitian ini digunakan metode *Levenshtein Distance*. Metode *Levenshtein Distance (LD)* digunakan untuk mengukur banyaknya perbedaan yang dimiliki dalam setiap kata, sehingga pada hasil akhirnya didapati persentase dari kemungkinan kesalahan ejaan kata. Dari hasil perhitungan metode *Levenshtein Distance* dengan contoh kata Pasia Menjadi Pasi di dapat jarak dengan nilai 1, maka kata Pasia dapat diubah menjadi Pasi. Dari hasil *User Acceptance Test (UAT)* didapatkanlah hasil persentase bahwa responden yang menjawab pertanyaan dengan jawaban Ya sebesar 84%, Tidak sebesar 6% dan Ragu-Ragu sebesar 10%. Dari hasil pengujian *Sugesstion Adequery* diperoleh akurasi sebesar 78,16% sehingga dapat disimpulkan bahwa aplikasi ini dapat diterima dengan baik.

**Kata kunci :** Bahasa Melayu Rokan hulu, *Levenshtein Distance*, *MySQL*, *PHP*

## **ABSTRACT**

*Rokan Hulu Malay is a language that has a mixture of languages, there are three language mixtures that form the basis of Hokan Hulu Malay, namely Mandailing, Malay and Minangkabau languages. Rokan Hulu Malay is used daily by the people of Rokan Hulu to communicate with each other both orally and in writing such as on social media, but many users on Rokan Hulu post and comment using the spelling of Malay words with incorrect spelling of words, such as abbreviations. words, writing errors are also called typos. This can lead to spelling errors in the meaning of the spelling of the Rokan Hulu Malay standard. In this research, the Levenshtein Distance method is used. The Levenshtein Distance (LD) method is used to measure the number of differences each word has, so that the final result is the percentage of possible word spelling errors. From the calculation results of the Levenshtein Distance method with the example of the word Pasia Becomes Pasi, the distance with a value of 1 is found, the word Pasia can be changed to Pasi. From the results of the User Acceptance Test (UAT) it was found that the percentage of respondents who answered questions with answers Yes was 84%, No was 6% and Doubtful was 10%. From the results of the Suggestion Adequery test, an accuracy of 78.16% was obtained so that it can be concluded that this application can be well received..*

**Keywords :** Malay Rokan Hulu, Levenshtein Distance, MySQL, PHP



## DAFTAR ISI

<b>PERSETUJUAN PEMBIMBING .....</b>	<b>i</b>
<b>PERSETUJUAN PENGUJI .....</b>	<b>ii</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<i>ABSTRACT .....</i>	<i>Error! Bookmark not defined.</i>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xv</b>
<b>DAFTAR SIMBOL .....</b>	<b>xvii</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Masalah .....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
<b>BAB 2 LANDASAN TEORI .....</b>	<b>7</b>
2.1 <i>Natural Language Processing (NLP)</i> .....	7

2.2	<i>Spelling Correction</i> .....	7
2.3	<i>Levenshtein Distance</i> .....	8
2.3.1	Operasi-Operasi pada <i>Levenshtein Distance</i> .....	8
2.3.2	Langkah-Langkah Algoritma <i>Levenshtein Distance</i> .....	10
2.4	<i>Suggestion Adequery</i> .....	11
2.5	Rokan Hulu.....	12
2.6	Bahasa Melayu Rokan Hulu.....	12
2.7	<i>Website</i> .....	14
2.8	<i>PHP (Hypertext Preprocessor)</i> .....	15
2.9	<i>HTML (Hyper Text Mark Up Language)</i> .....	15
2.10	<i>JavaScript</i> .....	15
2.11	<i>(CSS) Cascading Style Sheets</i> .....	16
2.12	Basis Data ( <i>Database</i> ).....	16
2.13	<i>MySQL (My Structure Query Language)</i> .....	17
2.14	<i>XAMPP</i> .....	17
2.15	<i>UML (Unified Modelling Language)</i> .....	17
2.15.1	<i>Use Case Diagram</i> .....	18
2.15.2	<i>Activity Diagram</i> .....	18
2.15.3	<i>Sequence Diagram</i> .....	18
2.15.4	<i>Class Diagram</i> .....	18

2.16	<i>Blackbox</i> .....	18
2.17	Penelitian Terkait .....	19
<b>BAB 3</b>	<b>METODOLOGI PENELITIAN</b> .....	<b>21</b>
3.1	Identifikasi Masalah .....	22
3.2	Analisa Masalah .....	22
3.3	Studi Literatur.....	22
3.4	Pengumpulan Data .....	23
3.5	Perancangan Sistem.....	23
3.6	Implementasi Sistem .....	24
3.7	Pengujian Sistem .....	24
3.8	Kesimpulan dan Saran.....	24
<b>BAB 4</b>	<b>ANALISA DAN PERANCANGAN</b> .....	<b>25</b>
4.1	Analisa Sistem.....	25
4.2	Analisa Kebutuhan Data.....	25
	4.2.2 Analisa Metode <i>Levenshtein Distance</i> .....	27
	4.2.3 Perhitungan Manual.....	28
4.3	Perancangan Sistem.....	46
	4.2.1 <i>Unified Modeling Language (UML)</i> .....	47
4.4	Perancangan Database .....	56
	4.4.1 <i>Database</i> .....	56

4.5	Perancangan Antarmuka.....	58
4.5.1	Halaman <i>login</i> .....	58
4.4.2	Halaman <i>Home</i> .....	59
4.4.3	Halaman Kamus.....	60
4.4.4	Halaman Koreksi ejaan.....	60
4.4.5	Halaman <i>User</i> .....	61
4.4.6	Halaman Kata .....	61
4.4.7	Halaman <i>User</i> .....	62
4.4.8	Halaman Koreksi Kata <i>User</i> .....	62
<b>BAB 5</b>	<b>IMPLEMENTASI DAN PENGUJIAN.....</b>	<b>63</b>
5.1	Implementasi .....	63
5.1.1	Lingkungan Implementasi.....	63
5.1.2	Batasan Implementasi .....	64
5.1.3	Hasil Implementasi.....	64
5.2	Pengujian .....	70
5.3	Pengujian <i>User Acceptance Test (UAT)</i> .....	72
5.4	Pengujian Aplikasi Terhadap Data.....	75
<b>BAB 6</b>	<b>PENUTUP.....</b>	<b>78</b>
6.1	Kesimpulan.....	78
6.2	Saran.....	79

<b>DAFTAR PUSTAKA .....</b>	<b>80</b>
-----------------------------	-----------

## DAFTAR TABEL

Tabel 2. 1 Penelitian terkait .....	19
Tabel 4.1 Variabel Contoh Data Kosa Kata Bahasa Melayu Rokan Hulu.....	27
Tabel 4.2 Matriks Sumber “pasia” Target “pasie” .....	28
Tabel 4.3 Matriks Hasil (1,1) .....	28
Tabel 4.4 Matriks Hasil (1,2) .....	29
Tabel 4.5 Matriks Hasil (1,3) .....	30
Tabel 4.6 Matriks Hasil (1,4) .....	31
Tabel 4.7 Hasil Matriks (1.5) .....	32
Tabel 4.8 Hasil Matriks (2.1) .....	32
Tabel 4.9 Hasil Matriks (2.2) .....	33
Tabel 4.10 Hasil Matriks (2.3) .....	34
Tabel 4.11 Hasil Matriks (2.4) .....	34
Tabel 4.12 Hasil Matriks (2.5) .....	35
Tabel 4.13 Hasil Matriks (3,1) .....	36
Tabel 4.14 Hasil Matriks (3,2) .....	37
Tabel 4.15 Hasil Matriks (3,3) .....	37
Tabel 4.16 Hasil Matriks (3,4) .....	38
Tabel 4.17 Hasil Matriks (3,5) .....	39
Tabel 4.18 Hasil Matriks (4,1) .....	39
Tabel 4.19 Hasil Matriks (4,2) .....	40
Tabel 4.20 Hasil Matriks (4,3) .....	41
Tabel 4.21 Hasil Matriks (4,4) .....	41
Tabel 4.22 Hasil Matriks (4,5) .....	42

Tabel 4.23 Hasil Matriks (5,1) .....	43
Tabel 4.24 Hasil Matriks (5,2) .....	44
Tabel 4.25 Hasil Matriks (5,3) .....	44
Tabel 4.26 Hasil Matriks (5,4) .....	45
Tabel 4.27 Hasil Matriks (5,5) .....	46
Tabel 4.28 Tabel Kata .....	57
Tabel 4.29 Tabel <i>User</i> .....	57
Tabel 4.30 Tabel <i>users</i> (Admin) .....	58
Tabel 5.1 Pengujia Dengan Akses <i>User</i> .....	70
Tabel 5.2 Pengujian Dengan Akses Admin .....	71
Tabel 5.3 Pengujian <i>User Acceptance Test (UAT)</i> .....	73
Tabel 5.4 Tabel Data Uji .....	75

## DAFTAR GAMBAR


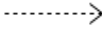







Gambar 2.1 Kumpulan (Trombo) Lafal Kata Bahasa Melayu Rokan Hulu .....	14
Gambar 3.1 <i>Flowchart</i> Metodologi Penelitian .....	21
Gambar 4.1 Kumpulan (Trombo) Lafal Kata Bahasa Melayu Rokan Hulu .....	26
Gambar 4.2 <i>Use Case Diagram</i> Aplikasi Koreksi Ejaan Kata Bahasa Melayu Rokan Hulu Menggunakan Metode <i>Levenshtein Distance</i> .....	47
Gambar 4.3 <i>Class Diagram</i> Aplikasi Koreksi Ejaan Kata Bahasa Melayu Rokan Hulu Menggunakan Metode <i>Levenshtein Distance</i> .....	48
Gambar 4.4 <i>Activity Diagram</i> Kamus .....	49
Gambar 4.5 <i>Activity Diagram</i> Koreksi Ejaan .....	50
Gambar 4.6 <i>Activity Diagram</i> User .....	50
Gambar 4.7 <i>Activity Diagram</i> Kata.....	50
Gambar 4.8 <i>Activity Diagram</i> Kata Dasar .....	51
Gambar 4.9 <i>Activity Diagram</i> Kata Dasar .....	51
Gambar 4.10 <i>Activity Diagram</i> Kata Dasar .....	52
Gambar 4.11 <i>Sequence Diagram</i> Login Admin.....	53
Gambar 4.12 <i>Sequence Diagram</i> Kamus .....	53
Gambar 4.13 <i>Sequence Diagram</i> Koreksi Ejaan.....	54
Gambar 4.14 <i>Sequence Diagram</i> User .....	54
Gambar 4.15 <i>Sequence Diagram</i> Kata.....	55
Gambar 4.16 <i>Sequence Diagram</i> Kata Dasar .....	55
Gambar 4.17 <i>Sequence Diagram</i> Ejaan Kata.....	56
Gambar 4.18 Halaman <i>Login</i> .....	59




Gambar 4.19 Halaman <i>Home</i> .....	59
Gambar 4.20 Halaman Kamus .....	60
Gambar 4.21 Halaman Koreksi Ejaan Kata .....	60
Gambar 4.22 Halaman <i>User</i> .....	61
Gambar 4.23 Halaman Kata.....	61
Gambar 4.24 Halaman <i>User</i> .....	62
Gambar 4.25 Halaman Koreksi Kata <i>User</i> .....	62



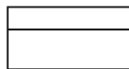

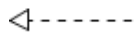
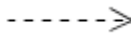

## DAFTAR SIMBOL

### 1. Use Case Diagram

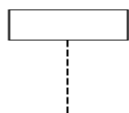
NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actor</i>	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan <i>use case</i> .
2		<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri ( <i>independent</i> ) akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri ( <i>independent</i> ).
3		<i>Generalization</i>	Hubungan dimana objek anak ( <i>descendent</i> ) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk ( <i>ancestor</i> ).
4		<i>Include</i>	Menspesifikasikan bahwa <i>use case</i> sumber secara <i>eksplisit</i> .
5		<i>Extend</i>	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan.
6		<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya.
7		<i>System</i>	Menspesifikasikan paket yang menampilkan sistem secara terbatas.
8		<i>Use Case</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor
9		<i>Collaboration</i>	Interaksi aturan-aturan dan elemen lain yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah dan elemen-elemennya ( <i>sinergi</i> ).


10		<i>Note</i>	Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi
----	---	-------------	---

## 2. Class Diagram






NO	GAMBAR	NAMA	KETERANGAN
1		<i>Generalization</i>	Hubungan dimana objek anak ( <i>descendent</i> ) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk ( <i>ancestor</i> ).
2		<i>Nary Association</i>	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek.
3		<i>Class</i>	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.
4		<i>Collaboration</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor
5		<i>Realization</i>	Operasi yang benar-benar dilakukan oleh suatu objek.
6		<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri ( <i>independent</i> ) akan memengaruhi elemen yang bergantung padanya elemen yang tidak mandiri
7		<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya

## 3. Sequence Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<i>LifeLine</i>	Objek <i>entity</i> , antarmuka yang saling berinteraksi.

2		<i>Message</i>	Spesifikasi dari komunikasi antar objek yang memuat informasi-informasi tentang aktifitas yang terjadi
3		<i>Message</i>	Spesifikasi dari komunikasi antar objek yang memuat informasi-informasi tentang aktifitas yang terjadi

#### 4. *Actifity Diagram*

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actifity</i>	Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain
2		<i>Action</i>	State dari sistem yang mencerminkan eksekusi dari suatu aksi
3		<i>Initial Node</i>	Bagaimana objek dibentuk atau diawali.
4		<i>Actifity Final Node</i>	Bagaimana objek dibentuk dan dihancurkan
5		<i>Fork Node</i>	Satu aliran yang pada tahap tertentu berubah menjadi beberapa aliran