

**PENERAPAN METODE *DAMEARAU LEVENSHTTEIN*  
*DISTANCE* UNTUK KOREKSI EJAAN BAHASA MELAYU  
TUANKU TAMBUSAI**

**(Studi Kasus Kecamatan Tambusai, Kabupaten Rokan Hulu, Riau)**

**SKRIPSI**

**OLEH:**

**NOVIA EKA REZA**

**NIM : 1837055**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PASIR PENGARAIAN**

**2022**

## PERSETUJUAN PEMBIMBING

### PENERAPAN METODE DAMERAU LEVENSHTTEIN DISTANCE UNTUK KOREK EJAAN BAHASA MELAYU TUANKU TAMBUSAI

---

Disetujui Oleh :

Pembimbing I



Luth Fimawahib, M. Kom  
NIDN. 101306891

Pembimbing II



Satria Riki Mustafa, M. Si  
NIDN. 10011039301

Diketahui Oleh :

Ketua Program Studi Teknik Informatika



Basrudin, S. Pd., M. Kom  
NIDN. 1020088702

**PERSETUJUAN PENGUJI**  
Skripsi ini telah diuji oleh  
**Tim Penguji Ujian Sarjana Komputer**  
**Program Studi Teknik Informatika**  
**Fakultas Ilmu Komputer**  
**Pada Tanggal 25 Juli 2022**

---

Tim Penguji :

- |   |            |   |
|---|------------|---|
| 1. <u>Luth Fimawahib, M. Kom</u><br>NIDN. 1009058707    | Ketua      | (  )   |
| 2. <u>Satria Riki Mustafa, M.Si</u><br>NIDN. 1001039301 | Sekretaris | (  )   |
| 3. <u>Erni Rouza, ST., M.Kom</u><br>NIDN. 1009058707    | Anggota    | (  )  |
| 4. <u>Budi Yanto, ST., M.Kom</u><br>NIDN. 1029058301    | Anggota    | (  ) |
| 5. <u>Basorudin, S.Pd., M.Kom</u><br>NIDN. 1020088702   | Anggota    | (  ) |

Mengetahui :

Dekan Fakultas Ilmu Komputer

Universitas Pasir Pengaraian



Hendri Maradana, M. Kom

NIDN. 1002038702

## LEMBAR PERNYATAAN

Dengan ini penulis menyatakan bahwa Skripsi yang berjudul “Penerapan Metode *Damerau Levenshtein Distance* untuk Koreksi Ejaan Bahasa Melayu Tuanku Tambusai” benar hasil penelitian penulis dengan arahan Dosen Pembimbing dan belum pernah diajukan dalam bentuk apapun untuk mendapatkan gelar Kesarjanaan. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis yang dicantumkan dalam daftar pustaka. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka penulis bersedia menerima sanksi berupa pencabutan gelar yang telah diperoleh karena Skripsi ini, serta lainnya sesuai norma yang berlaku di perguruan tinggi.

Pasir Pengaraian, 25 Juli 2022

Yang Membuat Pernyataan



**NOVIA EKA REZA**  
**NIM. 1837055**

## **KATA PENGANTAR**

*Assalammu'alaikum wa rahmatullah wa barokatuh*

*Alhamdulillah rabbil alamin*, segala puji syukur atas kehadiran Allah SWT yang senantiasa melimpahkan rahmat serta karunia-Nya, sehingga penulis mampu menyelesaikan Skripsi ini dengan baik. Shalawat beserta salam kita ucapkan kepada junjungan kita Rasulullah Muhammad SAW karena jasa beliau yang telah membawa manusia dari zaman kebodohan ke zaman yang penuh dengan ilmu pengetahuan seperti sekarang ini.

Skripsi ini disusun sebagai salah satu syarat untuk mendapatkan kelulusan pada jurusan Teknik Informatika Universitas Pasir Pengaraian. Banyak sekali pihak yang telah membantu penulis dalam menyusun 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 telah memberikan rahmat dan hidayahnya sehingga penulis mampu menyelesaikan skripsi ini dengan baik dan lancar.
2. Rasulullah SAW, yang telah membawa petunjuk bagi manusia agar menjadi manusia paling mulia derajatnya di sisi Allah SWT.
3. Kepada Omak, Etek, Acik, Mamak, abg, kakak, adik, dan keponakan yang selalu memberikan doa, motivasi, dan bantuan berupa materi serta selalu

memberikan sokongan untuk segera menyelesaikan Skripsi ini dengan baik dan tepat pada waktunya.

4. Bapak Dr. Hardianto, S.Pd., 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 ketua Prodi Teknik Informatika Universitas Pasir Pengaraian, sekaligus selaku penguji III yang telah membantu dan memberi arahan sekaligus masukan selama proses ujian.
7. Bapak Luth Fimawahib, M.Kom, selaku pembimbing I dan bapak Satria Riki Mustafa selaku pembimbing II yang sangat banyak membantu saya karena sudah meluangkan waktu dalam proses membimbing saya dalam pembuatan skripsi ini hingga tuntas, dan selalu memberikan motivasi kepada saya bahwa saya mampu untuk menyelesaikan skripsi ini.
8. Ibuk Erni Rouza, ST., M.Kom selaku penguji I, dan bapak Budi Yanto, ST., M.kom selaku penguji II saya yang sudah memberikan masukan dan arahan kepada saya
9. Teruntuk orang istimewa saya Frengki Hasibuan serta teman-teman terdekat saya, Desi Afriani, Merli Yanti, Neneng Daulay, Egha Havira, terimakasih karena tidak pernah bosan untuk memberikan semangat dan membantu saya dalam proses pembuatan skripsi ini.
10. Teruntuk diri sendiri, terimakasih diriku sudah bertahan di tengah rasa lelah dan capek ini, terimakasih sudah semangat padahal tubuh maunya

rebahan, terimakasih sudah berjuang dengan sisa kekuatan, ayo sedikit lagi mari kita tuntas kan ini bersama-sama hingga nanti apa yang sudah kita rencanakan akan tercapai.

11. Teman-teman seperjuangan saya di Program Studi Teknik Informatika, dan semua pihak yang tidak bisa saya sebutkan satu persatu, yang telah memberikan insiprasi dan semangat kepada penulis, selalu memberikan motivasi dan juga dukungan.

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

*Aamiin..*

*Wassalamua'alaikum wa rahmatullahi wa barakatuh*

Pasir Pengaraian, 25 Juli 2022

**NOVIA EKA REZA**  
**NIM. 1837055**

## ABSTRAK

Bahasa Melayu Tuanku Tambusai merupakan salah satu alat komunikasi yang digunakan oleh masyarakat di Kecamatan Tambusai. Namun sekarang ini, pengguna media sosial menggunakan berbagai bahasa antara lain, bahasa Indonesia, bahasa campuran dengan bahasa daerah, termasuk singkatan dan kata-kata gaul yang dapat menyebabkan salah dalam pengejaan. Untuk mengatasi permasalahan tersebut dibutuhkan suatu teknologi yang dapat mengoreksi kesalahan ejaan penulisan, salah satunya teknologi dalam bidang *Natural Language Processing (NLP)* dengan metode *Damerau Levenshtein Distance*. Metode *Damerau Levenshtein Distance* ini juga merupakan pengembangan dari metode *Levenshtein Distance* yang sebelumnya memiliki 3 operasi edit, dan pada metode *Damerau Levenshtein Distance* sudah memiliki 4 operasi edit dengan penambahan operasi *transposition*. Berdasarkan hasil pengujian menggunakan *suggestionadequacy*, maka didapatkan nilai 75,33% dari 300 kata uji, dimana 176 kata bernilai 1, 112 kata bernilai 0,5 dan 12 kata bernilai -0,5.

**Kata kunci** : Bahasa Melayu Tuanku Tambusai, *NLP*, *Damerau Levenshtein Distance*



## **ABSTRACT**

*Tuanku Tambusai Malay is one of the communication tools used by people in Tambusai sub-district. But nowadays, social media users use various languages, including Indonesian, a mixed language with regional languages, including abbreviations and slang words that can cause spelling mistakes. To overcome these problems, a technology that can correct spelling errors is needed, one of which is technology in the field of Natural Language Processing (NLP) with the Damerau Levenshtein Distance method. The Damerau Levenshtein Distance method is also a development of the Levenshtein Distance method which previously had 3 edit operations, and the Damerau Levenshtein Distance method already had 4 edit operations with the addition of a transposition operation. Based on the results of the test using the suggestion adequacy, a score of 75.33% was obtained from the 300 test words, of which 176 words were worth 1, 112 words were worth 0.5 and 12 words were worth -0.5.*

**Key words** : *Tuanku Tambusai Malay, NLP, Damerau Levenshtein Distance*

## DAFTAR ISI

	<b>Halaman</b>
<b>PERSETUJUAN PEMBIMBING .....</b>	<b>ii</b>
<b>PERSETUJUAN PENGUJI .....</b>	<b>iii</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>ABSTRAK.....</b>	<b>viii</b>
<b>ABSTRACT.....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xv</b>
<b>DAFTAR TABEL .....</b>	<b>xvi</b>
<b>DAFTAR SIMBOL.....</b>	<b>xviii</b>
<b>BAB I PENDAHULUAN</b>	
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	5
1.3. Batasan Masalah .....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian .....	6
1.6. Metodologi Penelitian .....	6
1.7.Sistematika Penulisan .....	7
<b>BAB II LANDASAN TEORI</b>	
2.1. <i>Artificial Intelligence</i> .....	9

2.2. <i>Natural Language Processing (NLP)</i> .....	9
2.3. <i>Text Preprocessing</i> .....	10
2.4. <i>Damerau Levenshtein Distance</i> .....	11
2.5. <i>Suggestion Adequacy</i> .....	13
2.6. Koreksi Ejaan .....	14
2.7. Bahasa Melayu .....	14
2.8. Bahasa Melayu Tuanku Tambusai .....	15
2.9. Aspek Morfologi Bahasa Melayu .....	16
2.10. <i>MySQL</i> .....	18
2.11. <i>PHP</i> .....	18
2.12. <i>XAMPP</i> .....	19
2.13. <i>Unified Modeling Language (UML)</i> .....	20
2.13.1. <i>Use Case Diagram</i> .....	21
2.13.2. <i>Diagram Aktifitas (Activity Diagram)</i> .....	21
2.13.3. <i>Diagram Urutan (Sequence Diagram)</i> .....	22
2.13.4. <i>Diagram Kelas (Class Diagram)</i> .....	22
2.15. Penelitian Terkait.....	22

### **BAB III METODOLOGI PENELITIAN**

3.1. Identifikasi.....	28
3.2. Perumusan masalah .....	29
3.3. Pengumpulan Data.....	29
3.4. Analisis.....	30
3.4.1. Analisis Metode <i>Damerau Levenshtein Distance</i> .....	31

3.4.2. Analisis Fungsi Sistem .....	31
3.5. Perancangan Sistem Aplikasi .....	31
3.6. Pembuatan Aplikasi .....	32
3.7. Implementasi Aplikasi .....	32
3.8. Pengujian Aplikasi.....	32
3.9. Kesimpulan Dan Saran .....	33

## **BAB IV ANALISIS DAN PERANCANGAN**

4.1. Analisis Sistem .....	34
4.1.1. Analisis Kebutuhan Data.....	34
4.2. Processing .....	37
4.3. Koreksi Ejaan .....	41
4.4 Perhitungan Manual <i>Danerau Levenshtein Distance</i> .....	42
4.4.1. Perhitungan Manual Kata Sanda.....	43
4.4.2. Perhitungan Manual Kata Atuk .....	58
4.4.3. Perhitungan Manual Kata Agih .....	60
4.4.4. Perhitungan Manual Kata Aku .....	61
4.5. Perancangan .....	62
4.6. Rancangan Sistem.....	63
4.6.1. <i>Use Case Diagram</i> .....	63
4.6.2. Diagram Aktifitas ( <i>Activity Diagram</i> ) .....	64
4.6.3. Diagram Urutan ( <i>Sequence Diagram</i> ).....	66
4.6.4. Diagram Kelas ( <i>Class Diagram</i> ) .....	68
4.7. Perancangan <i>Database</i> .....	69

4.7.1. Struktur Tabel .....	69
4.8. Perancangan Antar Muka.....	70
4.8.1. Halaman <i>Login</i> .....	70
4.8.2. Halaman Beranda.....	71
4.8.3. Halaman Data Kamus.....	71
4.8.4. Halaman Pengujian .....	72
4.8.5. Halaman Data Pengguna .....	72
<b>BAB V   IMPLEMENTASI DAN PENGUJIAN</b>	
1.1. Implementasi .....	73
1.1.1. Implementasi Perangkat Keras.....	73
1.1.2. Implementasi Perangkat Lunak.....	74
1.1.3. Implementasi Antar Muka Sistem.....	74
1.2. Pengujian Sistem .....	77
1.2.1. Pengujian <i>Login</i> .....	77
1.2.2. Pengujian Menu Beranda.....	78
1.2.3. Pengujian Menu Pengguna .....	79
1.3. Pengujian <i>Sugetion Adequacy</i> .....	79
1.4. Pengujian <i>UAT</i> .....	82
<b>BAB VI   PENUTUP</b>	
6.1. Kesimpulan.....	85
6.2. Saran .....	85
DAFTAR PUSTAKA .....	86
Lampiran 1 Kosakata Bahasa Melayu Tuanku Tambusai.....	89

Lampiran 2 Bukti <i>Chat Whatsapp</i> .....	116
Lampiran 3 Data Uji .....	117
Lampiran 4 Kuisisioner .....	119
Lampiran 5 Dokumentasi .....	129
RIWAYAT HIDUP .....	130

## DAFTAR GAMBAR

	Halaman
Gambar 2.1. Kamus Bahasa Melayu Tuanku Tambusai.....	16
Gambar 3.1. Tahapan Metodologi Penelitian.....	27
Gambar 4.1. Bukti <i>Chat Whatsapp</i> .....	36
Gambar 4.2. Data <i>Chat Whatsapp</i> .....	37
Gambar 4.3. <i>Flowchart</i> Analisis Sistem Koreksi Ejaan Bahasa Melayu Tuanku Tambusai.....	62
Gambar 4.4. <i>Use case diagram</i> .....	63
Gambar 4.5. <i>Activity Diagram Login</i> .....	64
Gambar 4.6. <i>Activity Diagram</i> Input .....	65
Gambar 4.7. <i>Activity Diagram</i> Pengujian .....	65
Gambar 4.8. <i>Activity Diagram</i> Hapus.....	66
Gambar 4.9. <i>Sequence Diagram Login</i> .....	67
Gambar 4.10. <i>Sequence Diagram</i> Kamus .....	67
Gambar 4.11. <i>Sequence Diagram</i> Pengujian.....	68
Gambar 4.12. <i>Sequence Diagram</i> Pengguna.....	68
Gambar 4.13. <i>Class Diagram</i> .....	69
Gambar 4.14. Tampilan Perancangan Halaman <i>Login</i> .....	70
Gambar 4.15. Tampilan Perancangan Halaman Beranda .....	71
Gambar 4.16. Tampilan Perancangan Halaman Kamus .....	71
Gambar 4.17. Tampilan Rancangan Halaman Pengujian .....	72
Gambar 4.18. Tampilan Rancangan Halaman Pengguna .....	72
Gambar 5.1. Tampilan Halaman <i>Login</i> .....	74
Gambar 5.2. Tampilan Halaman Beranda.....	75
Gambar 5.3. Tampilan Halaman Data Kamus.....	75
Gambar 5.4. Tampilan Halaman Pengujian .....	76
Gambar 5.5. Tampilan Halaman Pengguna .....	77

## DAFTAR TABEL



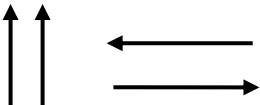
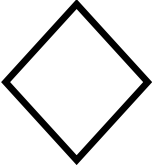




Tabel. 2.1. Proses Morfologi Melayu Tuanku Tambusai .....	17
Tabel. 2.2. Imbuhan Awal.....	17
Tabel. 2.3. Imbuhan Akhiran.....	17
Tabel. 2.4. Penelitian Terkait .....	23
Tabel 4.1. Variabel Contoh Data Kosakata.....	36
Tabel 4.2. Data <i>Whatsapp</i> .....	38
Tabel 4.3 <i>Cleaning</i> .....	38
Tabel 4.4. <i>Tokenizing</i> .....	39
Tabel 4.5. <i>Stemming</i> .....	40
Tabel 4.6. <i>Case Folding</i> .....	40
Tabel 4.7. Nilai Matriks D Kata Sanda.....	43
Tabel 4.8. Jarak D(1.1) Kata Sanda .....	44
Tabel 4.9. Jarak D(1.2)Kata Sanda.....	44
Tabel 4.10. Jarak D(1.3)Kata Sanda .....	45
Tabel 4.11. Jarak D(1.4)Kata Sanda .....	46
Tabel 4.12. Jarak D(1.5)Kata Sanda .....	46
Tabel 4.13. Jarak D(2.1)Kata Sanda .....	47
Tabel 4.14. Jarak D(2.2)Kata Sanda .....	48
Tabel 4.15. Jarak D(2.3)Kata Sanda .....	48
Tabel 4.16. Jarak D(2.4)Kata Sanda .....	49
Tabel 4.17. Jarak D(2.5)Kata Sanda .....	50
Tabel 4.18. Jarak D(3.1)Kata Sanda .....	50
Tabel 4.19. Jarak D(3.2)Kata Sanda .....	51
Tabel 4.20. Jarak D(3.3)Kata Sanda .....	51
Tabel 4.21. Jarak D(3.4)Kata Sanda .....	52
Tabel 4.22. Jarak D(3.5)Kata Sanda .....	53
Tabel 4.23. Jarak D(4.1)Kata Sanda .....	53
Tabel 4.24. Jarak D(4.2)Kata Sanda .....	54
Tabel 4.25. Jarak D(4.3)Kata Sanda .....	54



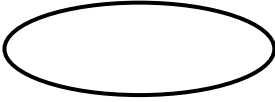

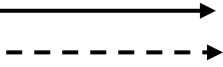


Tabel 4.26. Jarak D(4.4)Kata Sanda .....	55
Tabel 4.27. Jarak D(4.5)Kata Sanda .....	55
Tabel 4.28. Jarak D(5.1)Kata Sanda .....	56
Tabel 4.29. Jarak D(5.2)Kata Sanda .....	56
Tabel 4.30. Jarak D(5.3)Kata Sanda .....	57
Tabel 4.31. Jarak D(5.4)Kata Sanda .....	57
Tabel 4.32. Jarak D(5.5)Kata Sanda .....	58
Tabel 4.33. Jarak D(1.4) Kata Atuk.....	59
Tabel 4.34. Jarak D(5.4) Kata Atuk.....	59
Tabel 4.35. Jarak D(4.4) Kata Agih.....	60
Tabel 4.36. Jarak D(3.3) Kata Aku.....	61
Tabel 4.37. Kata Dasar.....	69
Tabel 4.38. Data Pengguna .....	70
Tabel 5.1. Pengujian <i>Login</i> .....	78
Tabel 5.2. Pengujian Menu Beranda.....	78
Tabel 5.3. Pengujian Pada Menu Pengguna.....	79
Tabel 5.4. Pemberian Nilai Pada Kata Tidak Normal .....	80
Tabel 5.5. Pertanyaan Pengujian UAT .....	82
Tabel 5.6. Indikator Bobot Nilai.....	82
Tabel 5.7. Hasil Jawaban Kuesioner dan Pengujian <i>UAT</i> .....	83
Tabel 5.8. Hasil Perhitungan <i>UAT</i> .....	83

## DAFTAR SIMBOL


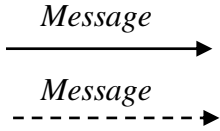
### 1. Simbol *Flowchart*

SIMBOL	NAMA	KETERANGAN
	Simbol komputerisasi	Menggambarkan proses yang dilakukan secara terkomputerisasi.
	<i>Input-Output</i>	Simbol yang menyatakan <i>input</i> dan <i>output</i> data.
	Simbol garis	Menggambarkan aliran proses dan dokumen.
	Simbol <i>decision</i> (Keputusan)	Menggambarkan proses pengambilan keputusan dalam sistem.
	<i>Terminator</i>	Untuk memulai ( <i>start</i> ) atau akhir ( <i>end</i> ) dari sesuatu kegiatan.
	<i>Document</i>	Mencetak keluaran dalam bentuk dokumen (melalui printer)
	<i>Preparation</i>	Simbol yang menyatakan penyediaan tempat penyimpanan suatu pengolahan untuk memberi harga awal
	<i>Display</i>	Simbol yang Menyatakan yang menyatakan peralatan output yang digunakan yaitu layar, plotter, printer dan sebagainya

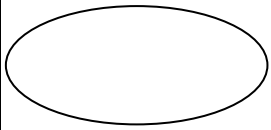
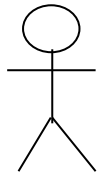
## 2. Simbol Activity Diagram

NO	Simbol	Nama	Deskripsi
1		<i>Action State</i>	Menggambarkan keadaan dari suatu elemen dalam suatu aliran aktifitas.
2		<i>State</i>	Menggambarkan kondisi suatu elemen.
3		<i>Control Flow</i>	Menggambarkan aliran aktifitas dari suatu elemen ke elemen yang lain.
4		<i>Initial State</i>	Menggambarkan titik awal siklus hidup suatu elemen.
5		<i>Final State</i>	Menggambarkan titik akhir yang menjadikan kondisi akhir suatu elemen.

## 3. Simbol Sequence Diagram

NO	Simbol	Nama	Deskripsi
1		<i>Object</i>	Menggambarkan pos-pos obyek yang mengirim dan penerima <i>message</i> .
2		<i>Message</i>	Menggambarkan aliran pesan yang dikirim oleh pos-pos obyek.

#### 4. Simbol Use Case

No	Simbol	Nama	Deskripsi
1		<i>Case</i>	Menggambarkan Proses/Kegiatan yang dapat dilakukan oleh aktor.
2		Aktor	Menggambarkan entitas/subyek yang dapat dilakukan suatu proses.
3	-End1      -End2	<i>Relation</i>	Relasi antara <i>case</i> dengan aktor ataupun <i>case dengan case</i> yang lain