

**DETEKSI KESALAHAN EJA KATA LULUH PADA PORTAL BERITA  
ONLINE BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA  
RABIN-KARP (STUDI KASUS: TRIBUNNEWS)**



Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Jason Tandiono**  
**00000035990**

UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2022**

**DETEKSI KESALAHAN EJA KATA LULUH PADA PORTAL BERITA  
ONLINE BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA  
RABIN-KARP (STUDI KASUS: TRIBUNNEWS)**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Jason Tandiono**  
**0000035990**

UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2022**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Jason Tandiono

Nomor Induk Mahasiswa : 00000035990

Program Studi : Informatika

Skripsi dengan judul:

**Deteksi Kesalahan Eja Kata Luluh Pada Portal Berita Online Berbahasa Indonesia Menggunakan Algoritma Rabin-Karp (Studi Kasus: Tribunnews)**

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 13 Juni 2023



(Jason Tandiono)

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

Skripsi dengan judul

### DETEKSI KESALAHAN EJA KATA LULUH PADA PORTAL BERITA ONLINE BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA RABIN-KARP (STUDI KASUS: TRIBUNNEWS)

oleh

Nama : Jason Tandiono  
NIM : 00000035990  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 27 Juni 2023

Pukul 15.00 s/d 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Farica Perdana Putri, S.Kom., M.Sc.)

NIDN: 0331019301

Penguji



(Dr. Ivrensa Zuhdi Pane, B.Eng., M.Eng.)

NIDN: 8812520016

Pembimbing



(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

Ketua Program Studi Informatika,



(Marlinda Vasty Overbeek, S.Kom., M.Kom.)

NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK  
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Jason Tandiono  
NIM : 00000035990  
Program Studi : Informatika  
Fakultas : Teknik dan Informatika  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**DETEKSI KESALAHAN EJA KATA LULUH PADA PORTAL BERITA  
ONLINE BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA  
RABIN-KARP (STUDI KASUS: TRIBUNNEWS)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 13 Juni 2023

Yang menyatakan

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

  
Jason Tandiono

**Halaman Persembahan / Motto**

"I never dreamed about success. I worked for it."

Estée Lauder



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Deteksi Kesalahan Eja Kata Luluh Pada Portal Berita Online Berbahasa Indonesia Menggunakan Algoritma Rabin-Karp (Studi Kasus: Tribunnews) dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.
6. Teman-teman saya yang telah membantu menyelesaikan tesis ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 13 Juni 2023

  
Jason Tandiono

**DETEKSI KESALAHAN EJA KATA LULUH PADA PORTAL BERITA  
ONLINE BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA  
RABIN-KARP (STUDI KASUS: TRIBUNNEWS)**

Jason Tandiono

**ABSTRAK**

Media sosial adalah media daring yang memberikan kesempatan bagi penggunanya untuk berpartisipasi, berbagi, dan menciptakan blog di dunia virtual dengan mudah. Portal berita daring adalah salah satu produk dari perkembangan teknologi internet di dunia jurnalisme. Pada setiap media pemberitaan, bahasa menjadi aspek utama dalam menyampaikan informasi. Namun, seiring berkembangnya teknologi, penerapan kaidah bahasa Indonesia yang baik dan benar mulai kurang diperhatikan. Beberapa kesalahan ejaan yang sering terjadi dalam penerapan kaidah bahasa Indonesia seperti, kesalahan penulisan huruf kapital, kesalahan penulisan huruf miring, kesalahan penulisan lambang bilangan, dan kesalahan penulisan tanda baca. Berdasarkan hal tersebut, penelitian ini dilakukan untuk membangun sebuah sistem yang dapat mendeteksi salah satu kesalahan penerapan bahasa Indonesia yang baik yaitu, kesalahan ejaan kata luluh dan memberikan solusi kata yang benar. Sistem akan dibuat menggunakan bahasa python dengan menerapkan algoritma Rabin-Karp. Algoritma Rabin-Karp bekerja dengan membandingkan nilai hash dari kata yang salah dengan kata yang ada pada *dataset*. Hasil penelitian yang dilakukan menunjukkan akurasi pendeteksian sistem sebesar 95,2% dimana nilai akurasi tersebut sama dengan nilai akurasi dari *stemmer* yang digunakan. Oleh karena itu, penelitian ini diharapkan dapat membantu mendeteksi kesalahan ejaan kata luluh dan memberikan solusi kata yang benar supaya kesalahan kaidah penulisan bahasa Indonesia dapat dihindari.

**Kata kunci:** Algoritma Rabin-Karp, Bahasa Indonesia, Kata Luluh, Kesalahan Ejaan, Portal Berita Daring

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



**DETECTION OF MELT WORD IN INDONESIAN LANGUAGE ONLINE  
NEWS PORTAL USING THE RABIN-KARP ALGORITHM (CASE  
STUDY: TRIBUNNEWS)**

Jason Tandiono

**ABSTRACT**

Social media is online media that provides an opportunity for its users to easily participate, share and create blogs in the virtual world. Online news portals are a product of the development of internet technology in the world of journalism. In every news media, language is the main aspect in conveying information. However, along with the development of technology, the application of good and correct Indonesian language rules is starting to get less attention. Some spelling errors that often occur in the application of Indonesian language rules, such as capital letter errors, italic writing errors, number symbol writing errors, and punctuation writing errors. Based on this, this research was conducted to build a system that can detect one of the mistakes in applying good Indonesian, namely, the misspelling of the word melt and provide the correct word solution. The system will be created using the Python language by applying the Rabin-Karp algorithm. The Rabin-Karp algorithm works by comparing the hash value of the wrong word with the word in dataset. The results of the research conducted show that the system detection accuracy is 95.2% where the accuracy value is the same as the accuracy value of the stemmer used. Therefore, this research is expected to help detect misspellings for the word luluh and provide correct word solutions so that mistakes in Indonesian writing rules can be avoided.

**Keywords:** Indonesian, Melt Word, Online News Portal, Rabin-Karp Algorithm, Spelling Error

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	vii
ABSTRACT . . . . .	viii
DAFTAR ISI . . . . .	ix
DAFTAR GAMBAR . . . . .	xi
DAFTAR TABEL . . . . .	xii
DAFTAR KODE . . . . .	xiii
DAFTAR LAMPIRAN . . . . .	xiv
BAB 1 PENDAHULUAN . . . . .	1
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Batasan Permasalahan . . . . .	3
1.4 Tujuan Penelitian . . . . .	3
1.5 Manfaat Penelitian . . . . .	3
1.6 Sistematika Penulisan . . . . .	4
BAB 2 LANDASAN TEORI . . . . .	5
2.1 Tinjauan Teori . . . . .	5
2.1.1 Portal Berita Daring . . . . .	5
2.1.2 Tribunnews . . . . .	5
2.1.3 Natural Language Processing . . . . .	5
2.1.4 Text Preprocessing . . . . .	6
2.1.5 Sastrawi . . . . .	7
2.1.6 Algoritma Rabin-Karp . . . . .	7
2.1.7 Fuzzy-Wuzzy . . . . .	8
2.1.8 Confusion Matrix . . . . .	9
BAB 3 METODOLOGI PENELITIAN . . . . .	12
3.1 Studi Literatur . . . . .	12
3.2 Pengumpulan Data . . . . .	12
3.3 Perancangan Sistem . . . . .	12
3.3.1 Diagram Alir . . . . .	12
3.4 Implementasi . . . . .	20
3.5 Pengujian dan Evaluasi Sistem . . . . .	21
3.6 Penulisan Laporan . . . . .	21
BAB 4 HASIL DAN DISKUSI . . . . .	22
4.1 Spesifikasi Sistem . . . . .	22
4.2 Implementasi Tampilan Antarmuka . . . . .	22
4.3 Implementasi Sistem Deteksi Kesalahan Ejaan Kata Luluh . . . . .	26
4.3.1 Pengumpulan <i>Dataset-Dataset</i> yang Dibutuhkan . . . . .	26
4.3.2 Implementasi Deteksi Kata Menggunakan Algoritma Rabin Karp . . . . .	27
4.4 Pengujian Sistem . . . . .	35
4.4.1 Pengujian Berdasarkan Jumlah Berita . . . . .	35
4.4.2 Pengujian Berdasarkan Nilai <i>Threshold/Cutoff</i> . . . . .	91

4.4.3	Pengujian Dengan Kluster . . . . .	97
4.5	Confusion Matrix . . . . .	98
4.5.1	Perhitungan Confusion Matrix 33 Berita . . . . .	99
4.5.2	Perhitungan Confusion Matrix 50 Berita . . . . .	99
4.5.3	Perhitungan Confusion Matrix 100 Berita . . . . .	100
BAB 5	SIMPULAN DAN SARAN . . . . .	102
5.1	Simpulan . . . . .	102
5.2	Saran . . . . .	102
DAFTAR PUSTAKA	. . . . .	104

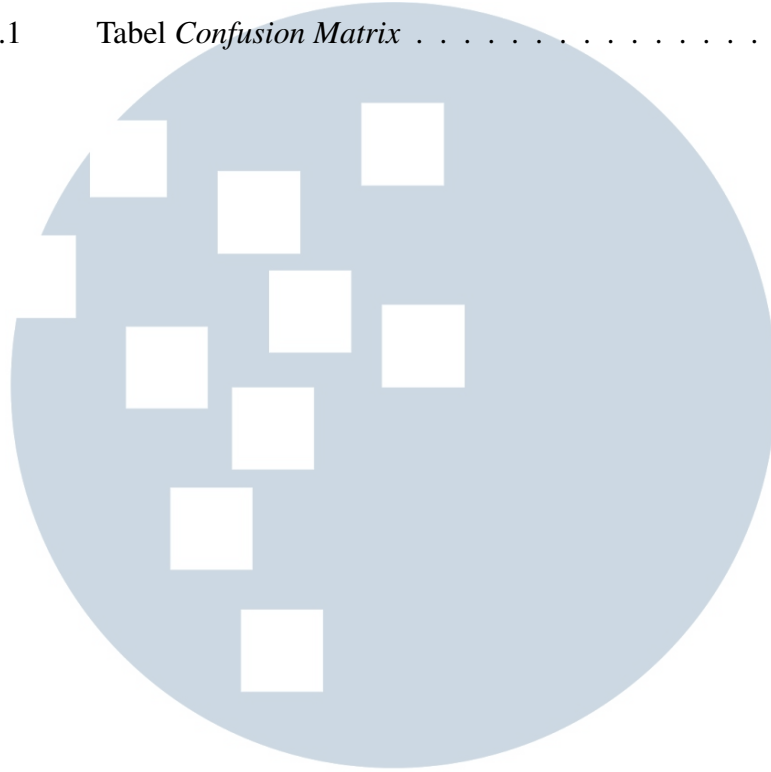


## DAFTAR GAMBAR

Gambar 3.1	Diagram Alir Utama . . . . .	13
Gambar 3.2	Diagram Alir Tahap <i>Text Preprocessing</i> . . . . .	14
Gambar 3.3	Diagram Alir Deteksi Kesalahan Ejaan Kata Luluh . . . . .	15
Gambar 3.4	Diagram Alir Algoritma Rabin Karp . . . . .	17
Gambar 3.5	Diagram Alir Perhitungan Nilai <i>Hash</i> . . . . .	18
Gambar 3.6	Diagram Alir Pencarian Solusi Kata Benar . . . . .	19
Gambar 4.1	Tampilan halaman indeks . . . . .	22
Gambar 4.2	Tampilan halaman hasil jika tidak ditemukan kesalahan ejaan kata luluh . . . . .	23
Gambar 4.3	Tampilan halaman hasil jika ditemukan kesalahan ejaan kata luluh dan solusi katanya . . . . .	24
Gambar 4.4	Tampilan halaman hasil jika ditemukan kesalahan ejaan kata luluh dan tidak ditemukan solusi katanya . . . . .	25
Gambar 4.5	Tampilan <i>dataset</i> kata benar . . . . .	26
Gambar 4.6	Tampilan <i>dataset</i> kata dasar . . . . .	26
Gambar 4.7	Tampilan <i>dataset</i> berita . . . . .	27
Gambar 4.8	Tampilan halaman indeks saat menguji 17 berita . . . . .	36
Gambar 4.9	Tampilan hasil pengujian 17 berita . . . . .	36
Gambar 4.10	Tampilan halaman indeks saat menguji 33 berita . . . . .	46
Gambar 4.11	Tampilan hasil pengujian 33 berita . . . . .	46
Gambar 4.12	Tampilan halaman indeks saat menguji 50 berita . . . . .	65
Gambar 4.13	Tampilan hasil pengujian 50 berita . . . . .	65
Gambar 4.14	Tampilan halaman indeks pengujian dengan nilai <i>threshold</i> 88 . . . . .	92
Gambar 4.15	Tampilan halaman hasil pengujian dengan nilai <i>threshold</i> 88 . . . . .	92
Gambar 4.16	Tampilan halaman indeks pengujian dengan nilai <i>threshold</i> 80 . . . . .	94
Gambar 4.17	Tampilan halaman hasil pengujian dengan nilai <i>threshold</i> 80 . . . . .	94
Gambar 4.18	Tampilan halaman indeks pengujian dengan nilai <i>threshold</i> 70 . . . . .	95
Gambar 4.19	Tampilan halaman hasil pengujian dengan nilai <i>threshold</i> 70 . . . . .	96
Gambar 4.20	Tampilan halaman indeks pengujian dengan kluster . . . . .	97
Gambar 4.21	Tampilan halaman hasil pengujian dengan kluster . . . . .	98
Gambar 4.22	<i>Confusion matrix</i> 33 berita . . . . .	99
Gambar 4.23	<i>Confusion matrix</i> 50 berita . . . . .	99
Gambar 4.24	<i>Confusion matrix</i> 100 berita . . . . .	100

## DAFTAR TABEL

Tabel 2.1	Tabel <i>Confusion Matrix</i> . . . . .	9
-----------	---	---



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

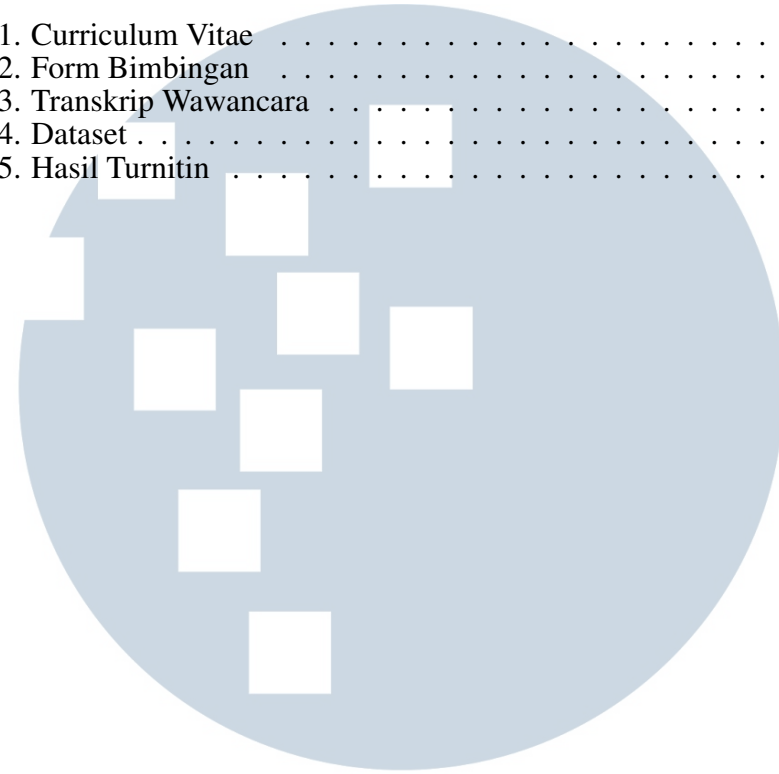
## DAFTAR KODE

2.1	Pseudocode Algoritma Levenshtein Distance . . . . .	8
4.1	Preprocessing Data . . . . .	28
4.2	Deteksi Kesalahan Ejaan Kata Luluh . . . . .	28
4.3	Fungsi Perhitungan Hash . . . . .	32
4.4	Fungsi Algoritma Rabin-Karp . . . . .	32
4.5	Fungsi Solusi Kata Benar . . . . .	33



## DAFTAR LAMPIRAN

Lampiran 1. Curriculum Vitae . . . . .	106
Lampiran 2. Form Bimbingan . . . . .	108
Lampiran 3. Transkrip Wawancara . . . . .	110
Lampiran 4. Dataset . . . . .	121
Lampiran 5. Hasil Turnitin . . . . .	122



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA