

**DETEKSI DAN PERBAIKAN PENULISAN KATA BAKU BAHASA  
INDONESIA PADA BERITA DARING DENGAN ALGORITMA  
LEVENSHTTEIN DISTANCE**



**SKRIPSI**

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

**Jezreel Kosasih**  
**00000043909**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2024**

**DETEKSI DAN PERBAIKAN PENULISAN KATA BAKU BAHASA  
INDONESIA PADA BERITA DARING DENGAN ALGORITMA  
LEVENSHTTEIN DISTANCE**



**SKRIPSI**

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

**Jezreel Kosasih**

**00000043909**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

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

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Jezreel Kosasih

NIM : 00000043909

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Tesis/Skripsi/Tugas Akhir/Laporan Magang/MBKM saya yang berjudul:

**Deteksi dan Perbaikan Penulisan Kata Baku Bahasa Indonesia pada Berita Daring dengan Algoritma Levenshtein Distance**

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 13 Mei 2024



(Jezreel Kosasih)

UMN  
UNIVERSITA  
MULTIMEDIA  
NUSANTARA

**HALAMAN PENGESAHAN**

Skripsi dengan judul

**DETEKSI DAN PERBAIKAN PENULISAN KATA BAKU BAHASA  
INDONESIA PADA BERITA DARING DENGAN ALGORITMA  
LEVENSHTEIN DISTANCE**

oleh

Nama : Jezreel Kosasih  
NIM : 00000043909  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 30 Mei 2024

Pukul 13.00 s/s 15.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

Ketua Sidang



(David Agustriawan, S.Kom., M.Sc.,  
Ph.D.)

NIDN: 0525088601

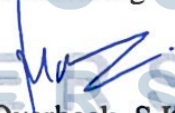
Penguji



(Samiaji Bintang Nusantara S.T., M.A.)

NIDN: 0324037702

Pembimbing



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

NIDN: 0818038501

Dr. Ketua Program Studi Informatika,



(Dr. Eng. Niki Prastomo, S.T., M.Sc.)

NIDN: 0419128203

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Jezreel Kosasih  
NIM : 00000043909  
Program Studi : Informatika  
Jenjang : S1  
Judul Karya Ilmiah : Deteksi dan Perbaikan Penulisan Kata Baku  
Bahasa Indonesia pada Berita Daring dengan  
Algoritma Levenshtein Distance

Menyatakan dengan sesungguhnya bahwa saya bersedia:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial. Saya tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

Tangerang, 09 Juni 2024

()

\* Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

**Halaman Persembahan / Motto**

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)

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 dan Perbaikan Penulisan Kata Baku Bahasa Indonesia pada Berita Daring dengan Algoritma Levenshtein Distance 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:

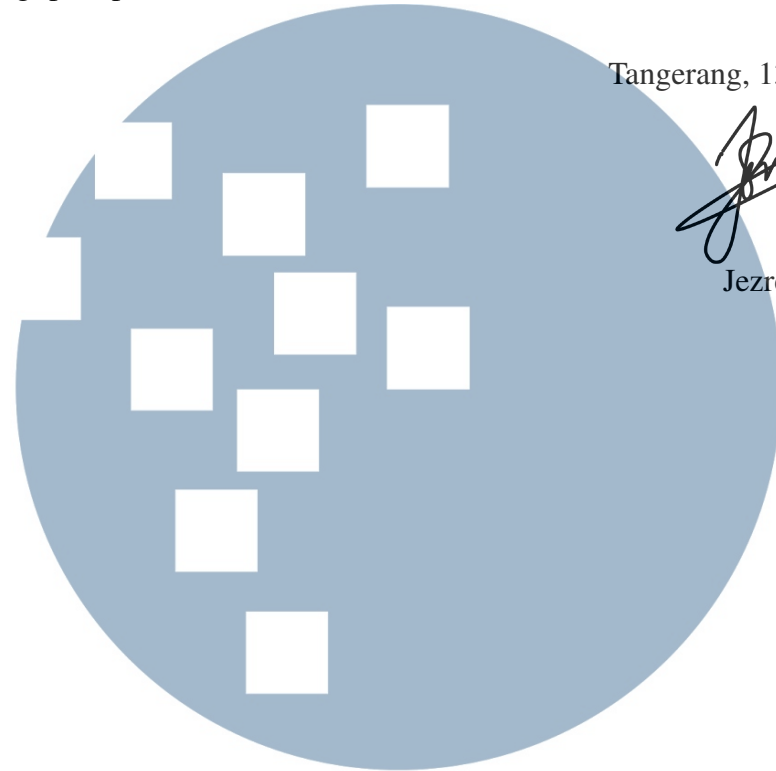
1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak 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. Ibu Dr. Niknik Mediyawati, S.Pd, M.Hum, Bapak Bintang Samiaji Nusantara, S.T, M.A. sebagai Dosen yang telah banyak membantu dan memberikan bimbingan atas terselesainya Skripsi ini.
6. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan Skripsi ini.
7. Teman-teman saya Brigitta Nissya, Jessica Sheridan, Reyhan Phillies Wijaya, Indiana Jane, Cristophrous Tritama, Claudia Nicola, Charina Eliani, Clarissa Liani, Delicia Karen, dan Farion Wang yang telah mendukung dan menemani saya selama proses penyelesaian Skripsi ini
8. Teman-teman seperjuangan U-Tapis yaitu Vallencius Gavriel, Antonius Kevin B. S., Kevin Gleneagles, dan Jonathan Octavien yang telah belajar dan berjuang bersama selama proses penyelesaian Skripsi ini.

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

Tangerang, 13 Mei 2024



Jezreel Kosasih



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



**DETEKSI DAN PERBAIKAN PENULISAN KATA BAKU BAHASA  
INDONESIA PADA BERITA DARING DENGAN ALGORITMA  
LEVENSHTEIN DISTANCE**

Jezreel Kosasih

**ABSTRAK**

Penelitian deteksi dan perbaikan kesalahan kata baku didasari oleh kemajuan teknologi dan *artificial intelligence* serta masalah kesalahan yang umum terjadi dalam penulisan artikel jurnalistik akibat keterbatasan waktu dan banyaknya tuntutan kerja. Penelitian deteksi dan perbaikan kesalahan kata baku turut didasari oleh penelitian sebelumnya yaitu penelitian U-Tapis. Penelitian merupakan pengembangan penelitian U-Tapis untuk menciptakan sebuah aplikasi berbasis *machine learning* untuk melakukan deteksi dan perbaikan kesalahan kata atau tata kalimat dalam bahasa Indonesia. Penelitian menggunakan algoritma Levenshtein Distance sebagai algoritma utama untuk memproses *input* artikel. Algoritma ini akan mencocokkan tiap karakter pada sebuah kata dan menghasilkan nilai *ratio* atau *similarity*. Penelitian menunjukkan keberhasilan dengan akurasi deteksi kata tidak baku hingga 100% dan F1 score mencapai 1.0. Penelitian terbukti efektif dalam mendeteksi kata tidak baku pada artikel berita dan mampu menggantikan kata tersebut dengan kata yang baku.

**Kata kunci:** Levenshtein Distance, *ratio*, U-Tapis, jurnalistik



# **DETECTION AND IMPROVEMENT OF INDONESIAN STANDARDIZED WORD WRITING IN ONLINE NEWS DRAFT WITH LEVENSHTTEIN DISTANCE ALGORITHM**

Jezreel Kosasih

## **ABSTRACT**

*The research on standardized word error detection and correction is based on technological advances and artificial intelligence as well as the problem of errors that are common in writing journalistic articles due to time constraints and work demands. The research on standardized word error detection and correction is also based on previous research, namely U-Tapis research. The research is a development of U-Tapis research to create a text-based application machine learning to detect and correct word or sentence errors in Indonesian. The research uses the Levenshtein Distance algorithm as the main algorithm for processing input articles. This algorithm will match each character in a word and produce a ratio or similarity value. The research shows success with the accuracy of nonstandard word detection up to 100% and F1 score reaching 1.0. The research proved to be effective in detecting nonstandard words in news articles and was able to replace the word with a standard word.*

**Kata kunci:** *Levenshtein Distance, ratio, U-Tapis, journalistic*



## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PERSETUJUAN . . . . .	iii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	viii
ABSTRACT . . . . .	ix
DAFTAR ISI . . . . .	x
DAFTAR GAMBAR . . . . .	xii
DAFTAR TABEL . . . . .	xiii
DAFTAR KODE . . . . .	xiv
DAFTAR LAMPIRAN . . . . .	xv
<b>BAB 1 PENDAHULUAN . . . . .</b>	<b>1</b>
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Batasan Masalah . . . . .	3
1.4 Tujuan Penelitian . . . . .	3
1.5 Manfaat Penelitian . . . . .	4
1.6 Sistematika Penulisan . . . . .	4
1.6.1 BAB 1 PENDAHULUAN . . . . .	4
1.6.2 BAB 2 LANDASAN TEORI . . . . .	4
1.6.3 BAB 3 METODOLOGI PENELITIAN . . . . .	5
1.6.4 BAB 4 HASIL DAN DISKUSI . . . . .	5
1.6.5 BAB 5 SIMPULAN DAN SARAN . . . . .	5
<b>BAB 2 LANDASAN TEORI . . . . .</b>	<b>6</b>
2.1 Bahasa Indonesia . . . . .	6
2.2 Kata Baku . . . . .	6
2.3 Natural Language Processing . . . . .	7
2.4 Text Preprocessing . . . . .	7
2.5 Levenshtein Distance . . . . .	9
2.6 Pustaka FuzzyWuzzy . . . . .	10
<b>BAB 3 METODOLOGI PENELITIAN . . . . .</b>	<b>15</b>
3.1 Studi Literatur . . . . .	17
3.2 Pengumpulan Data . . . . .	17
3.3 <i>Pre-processing</i> Teks Berita Daring . . . . .	18
3.4 Algoritma Levenshtein Distance . . . . .	18
3.5 Implementasi Algoritma . . . . .	21
3.6 Pengembangan Situs Web . . . . .	21
3.7 Uji Coba dan Evaluasi . . . . .	23
<b>BAB 4 HASIL DAN DISKUSI . . . . .</b>	<b>25</b>
4.1 Spesifikasi Sistem . . . . .	25
4.2 Dataset . . . . .	25
4.2.1 Split Dataset . . . . .	26
4.3 Pre-processing . . . . .	27
4.4 Permodelan . . . . .	28

4.4.1	Pencocokan string dengan algoritma Levenshtein Distance dari pustaka fuzzywuzzy . . . . .	28
4.4.2	Proses menampilkan <i>output</i> program . . . . .	30
4.4.3	Fitur tambahan . . . . .	32
4.5	Aplikasi Web . . . . .	32
4.6	Diagram Alir Aplikasi Web . . . . .	35
4.7	Berkas Penyusun Aplikasi Web . . . . .	38
4.7.1	Berkas program.py . . . . .	38
4.7.2	Berkas template.html . . . . .	41
4.8	Evaluasi Program . . . . .	43
4.8.1	Akurasi . . . . .	48
4.8.2	F1 Score . . . . .	49
4.8.3	Mean Squared Error Score . . . . .	50
BAB 5	SIMPULAN DAN SARAN . . . . .	52
5.1	Simpulan . . . . .	52
5.2	Saran . . . . .	52
	DAFTAR PUSTAKA . . . . .	54



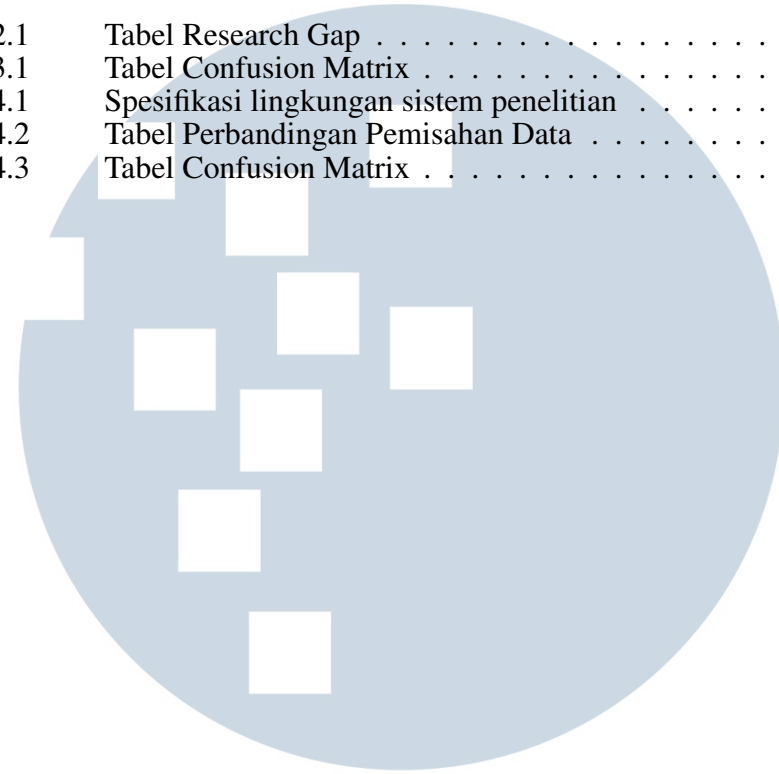
## DAFTAR GAMBAR

Gambar 2.1	Hasil deteksi pada halaman situs penelitian U-Tapis kata baku . . . . .	12
Gambar 2.2	Hasil deteksi pada halaman situs ejaan.id . . . . .	12
Gambar 2.3	Hasil deteksi pada halaman situs typhoonline . . . . .	13
Gambar 2.4	Hasil deteksi pada halaman situs Gramatika . . . . .	13
Gambar 2.5	Hasil deteksi pada halaman situs cek-ejaan.com . . . . .	14
Gambar 3.1	Diagram metode penelitian . . . . .	16
Gambar 3.2	Diagram End to End Process . . . . .	17
Gambar 3.3	Proses <i>pre-processing</i> input teks berita . . . . .	18
Gambar 3.4	Diagram alir Algoritma Levenshtein Distance . . . . .	20
Gambar 3.5	Rancangan tampak halaman web untuk <i>input</i> teks . . . . .	22
Gambar 3.6	Rancangan tampak halaman web untuk unggah berkas . . . . .	22
Gambar 4.1	Tampak halaman web <i>input</i> teks . . . . .	33
Gambar 4.2	Tampak halaman web unggah berkas . . . . .	33
Gambar 4.3	<i>Output</i> dari <i>input</i> teks . . . . .	34
Gambar 4.4	<i>Output</i> dari <i>input</i> berkas txt bagian 1 . . . . .	34
Gambar 4.5	<i>Output</i> dari <i>input</i> berkas txt bagian 2 . . . . .	34
Gambar 4.6	Diagram alir utama aplikasi web . . . . .	35
Gambar 4.7	Diagram alir fungsi <i>replaced_words_in_sentences</i> . . . . .	36
Gambar 4.8	Diagram alir fungsi <i>suggest_correct_words</i> . . . . .	37
Gambar 4.9	Hasil evaluasi program menggunakan pustaka <i>sklearn</i> python . . . . .	44
Gambar 4.10	Hasil 1 <i>false detection</i> pada program . . . . .	44
Gambar 4.11	Hasil 2 <i>false detection</i> pada program . . . . .	45
Gambar 4.12	Hasil 3 <i>false detection</i> pada program . . . . .	45
Gambar 4.13	Hasil 4 <i>false detection</i> pada program . . . . .	45
Gambar 4.14	Hasil benar 1 deteksi kata tidak baku pada program . . . . .	46
Gambar 4.15	Hasil benar 2 deteksi kata tidak baku pada program . . . . .	46
Gambar 4.16	Hasil benar 3 deteksi kata tidak baku pada program . . . . .	46
Gambar 4.17	Hasil benar 4 deteksi kata tidak baku pada program . . . . .	47
Gambar 4.18	Hasil benar 5 deteksi kata tidak baku pada program . . . . .	47
Gambar 4.19	Hasil benar 6 deteksi kata tidak baku pada program . . . . .	47
Gambar 4.20	Grafik Learning Curves Model . . . . .	48
Gambar 4.21	Mean Squared Error dari Model Deteksi dan Koreksi Kata Baku . . . . .	51

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR TABEL

Tabel 2.1	Tabel Research Gap . . . . .	11
Tabel 3.1	Tabel Confusion Matrix . . . . .	23
Tabel 4.1	Spesifikasi lingkungan sistem penelitian . . . . .	25
Tabel 4.2	Tabel Perbandingan Pemisahan Data . . . . .	26
Tabel 4.3	Tabel Confusion Matrix . . . . .	48



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR KODE

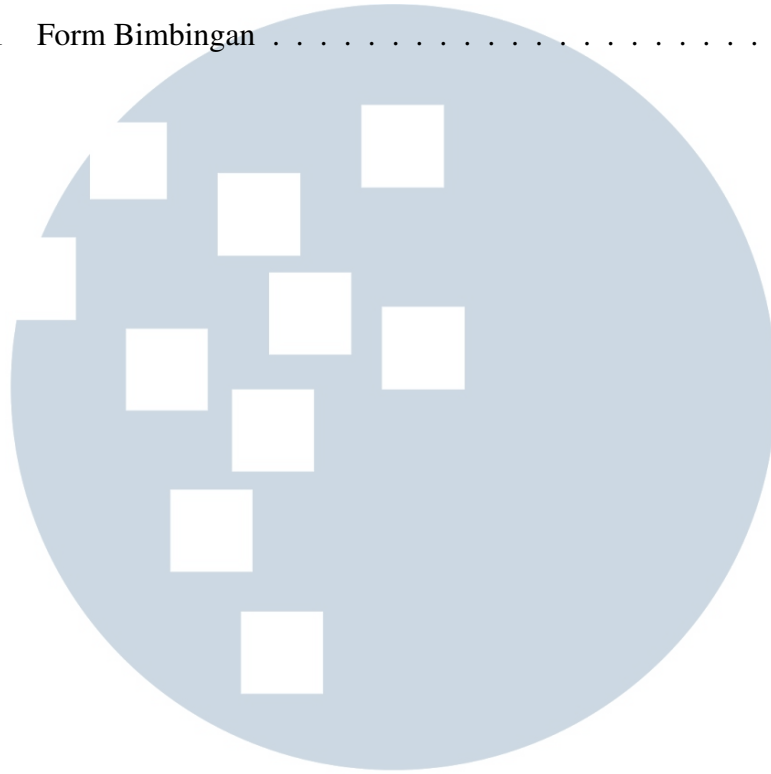
4.1	Proses <i>pre-processing</i> data <i>input</i> dari <i>user</i> . . . . .	27
4.2	Fungsi <i>ratio</i> pada pustaka <i>fuzzywuzzy</i> . . . . .	28
4.3	Fungsi <i>ratio</i> pada objek <i>SequenceMatcher</i> . . . . .	29
4.4	Implementasi metode <i>ratio</i> pada program <i>U-Tapis kata baku</i> . . . . .	30
4.5	Proses menampilkan <i>output</i> . . . . .	30
4.6	Pendeteksian kutipan langsung . . . . .	32
4.7	Fungsi <i>replace_words_in_sentence</i> . . . . .	38
4.8	Fungsi <i>suggest_correct_words</i> . . . . .	39
4.9	Potongan kode fungsi <i>main</i> untuk mengirimkan <i>output</i> ke tampilan web . . . . .	40
4.10	potongan kode <i>JavaScript</i> pada berkas <i>template.html</i> . . . . .	41





**DAFTAR LAMPIRAN**

Lampiran 1 Form Bimbingan . . . . . 57



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA