

**IMPLEMENTASI AUTOCORRECT DENGAN METODE
N-GRAM DAN LEVENSHTTEIN DISTANCE PADA APLIKASI
CHAT DENGAN FITUR PENERJEMAHAN**

SKRIPSI



Disusun Oleh:

Muhammad Nawwaf Naufal

210602043

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

2025

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa, berkat rahmat dan nikmat yang telah Dia berikan kepada penulis sehingga dapat menyelesaikan proposal skripsi dengan judul **“IMPLEMENTASI AUTOCORRECT DENGAN METODE N-GRAM DAN LEVENSHTAIN DISTANCE PADA APLIKASI CHAT DENGAN FITUR PENERJEMAHAN”** dengan lancar. Skripsi ini digunakan sebagai persyaratan menjadi Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Gresik.

Selesainya Skripsi ini tentunya tidak lepas dari bimbingan, arahan, dukungan, saran serta doa dari banyak pihak. Dikarenakan itu, pada kesempatan kali ini penulis akan menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT, yang senantiasa memberikan kesehatan, kesabaran, dan ketabahan pada penulis.
2. Kedua orang tua dan seluruh keluarga
3. Ucapan terima kasih kepada Ibu Umi Chotijah, S.Kom., M.Kom dan Bapak Harunur Rosyid, S.T., M.Kom. selaku dosen pembimbing.
4. Rekan-rekan seperjuangan Teknik Informatika Angkatan 2021 yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata kesempurnaan. Oleh karena itu penulis mengharapkan kritik dan saran yang sifatnya membangun bagi penulis, pembaca, dan masyarakat sehingga laporan ini dapat bermanfaat.

Lamongan, 24 Juni 2025

Penulis

Muhammad Nawwaf Naufal

DAFTAR ISI

LEMBAR PERSETUJUAN SKRIPSI	i
LEMBAR PENGESAHAN SKRIPSI	ii
KATA PENGANTAR	iii
LEMBAR PERNYATAAN KEASLIAN SKRIPSI	iv
DAFTAR ISI	iv
DAFTAR GAMBAR	viii
DAFTAR TABLE	ix
DAFTAR PERSAMAAN	xi
ABSTRAK.....	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
2.1 Rumusan Masalah.....	2
3.1 Tujuan Penelitian.....	2
4.1 Batasan Masalah	3
5.1 Manfaat Penelitian.....	3
6.1 Metodologi Penelitian.....	3
7.1 Sistematika Penulisan	4
BAB II LANDASAN TEORI	6
2.1 Penerjemah Otomatis	6
2.2 Text Processing.....	6
2.3 N-Gram.....	7
2.4 Levenshtein Distance	8

2.5 Probabilitas N-Gram.....	9
2.6 Pengukuran Akurasi	11
2.7 Penelitian Terdahulu	11
BAB III ANALISA DAN PERANCANGAN SISTEM	16
3.1 Analisa System	16
3.2 Hasil Analisa System.....	16
3.2.1 Pembentukan Unigram dan Bigram.....	17
3.2.2 Flowchart Proses Pengolahan Pesan	18
3.2.3 Flowchart Pengolahan Data	19
3.2.4 Flowchart Algoritma N-Gram Data Latih.....	21
3.2.5 Flowchart Levenshtein Distance dan N-Gram pada Autocorrect	22
3.3 Representasi Model	24
3.3.1 Perhitungan Frekuensi	24
3.3.2 Perhitungan Probabilitas Menggunakan Metode Maximum Likelihood Estimation (MLE) dan Laplace Smoothing	27
3.3.2.1 Perhitungan Probabilitas Unigram.....	27
3.3.2.2 Perhitungan Probabilitas Bigram	27
3.3.3 Perhitungan Menggunakan Levenshtein Distance	30
3.3.3.1 Definisi dan Matriks Awal	30
3.3.3.2 Hasil Perhitungan Levenshtein Distance	31
3.3.4 Perhitungan Skor	31
3.3.5 Pengujian dan Penilaian Kinerja Sistem	32
3.3.5.1 Perhitungan Akurasi	32
3.3.5.2 Perhitungan Akurasi Rata-rata	33
3.4 Perancangan Sistem	33
3.4.1 Diagram Konteks	33

3.4.2 Diagram Berjenjang.....	34
3.4.3 Data Flow Diagram (DFD) Level 1.....	36
3.4.4 Perancangan Basis Data	37
3.4.5 Perancangan Antar Muka.....	40
3.4.6 Spesifikasi Pengembangan System	46
3.4.7 Perencanaan Pengujian System	47
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM	49
4.1 Implementasi Sistem.....	49
4.1.1 Preprocessing dan Pembentukan Unigram serta Bigram	49
4.1.2 Perhitungan Probabilitas.....	50
4.1.3 Fungsi Autocorrect menggunakan Levenshtein Distance.....	50
4.1.4 Pencarian Kandidat Koreksi.....	51
4.1.5 Fungsi Autocorrect.....	52
4.1.6 Pengelolaan Kamus dan Dataset.....	53
4.1.7 Implementasi Penerjemahan.....	54
4.2 Pengujian Sistem.....	55
4.2.1 Halaman Login	55
4.2.2 Halaman Registrasi.....	56
4.2.3 Halaman Utama	56
4.2.4 Halaman Contacts	57
4.3 Analisa Pengujian Sistem	58
4.3.1 Skenario Pengujian Sistem	58
4.3.2 Akurasi Pengujian Sistem	79
BAB V PENUTUP	81
5.1 Kesimpulan.....	81
5.2 Saran	81

DAFTAR PUSTAKA	82
Lampiran	85



DAFTAR GAMBAR

Gambar 3.1. Flowchart Proses Pengolahan Pesan.....	18
Gambar 3.2. Flowchart Pengolahan Data Autocorrect	20
Gambar 3.3 Proses Ekstraksi Data Menjadi N-Gram untuk Autocorrect.....	21
Gambar 3.5 Proses Levenshtein Distance pada Autocorrect	23
Gambar 3.6. Diagram Konteks Sistem Chat dengan Integrasi N-Gram dan Levenshtein Distance	34
Gambar 3.7. Diagram Berjenjang Sistem Chat dengan Integrasi N-Gram	34
Gambar 3.8. Data Flow Diagram (DFD) Level 1 Sistem Chat dengan Integrasi N-Gram.....	36
Gambar 3.9. Diagram ERD App Chat	38
Gambar 3.10. Halaman Login.....	41
Gambar 3.11. Halaman Registrasi	42
Gambar 3.12. Halaman Verifikasi Email.....	42
Gambar 3.13. Halaman Verifikasi OTP	43
Gambar 3.14. Halaman Atur Ulang Sandi.....	44
Gambar 3.15. Halaman Utama	45
Gambar 3.16. Halaman Contact	45
Gambar 3.17. Halaman Profile	46
Gambar 4.1. Halaman Login	56
Gambar 4.2. Halaman Registrasi.....	56
Gambar 4.3. Halaman Utama	57
Gambar 4.4. Halaman Contacts	58

DAFTAR TABLE

Tabel 3.1 Data Latih Bahasa Indonesia <i>Unigram</i>	24
Tabel 3.2 Data Latih Bahasa Indonesia <i>Bigram</i>	25
Tabel 3.3 Data Latih Bahasa Inggris <i>Unigram</i>	25
Tabel 3.4 Data Latih Bahasa Inggris <i>Bigram</i>	26
Tabel 3.5 Data Latih Spanyol <i>Unigram</i>	26
Tabel 3.6 Data Latih Spanyol <i>Bigram</i>	26
Tabel 3.7 Hasil Sorting	27
Tabel 3.8 Hasil Perhitungan Probabilitas <i>Unigram</i> Dataset Bahasa Indonesia	28
Tabel 3.9 Hasil Perhitungan Probabilitas <i>Bigram</i> Dataset Bahasa Indonesia	29
Tabel 3.10 Hasil Perhitungan Probabilitas <i>Unigram</i> Dataset Bahasa Inggris	29
Tabel 3.11 Hasil Perhitungan Probabilitas <i>Bigram</i> Dataset Bahasa Inggris	29
Tabel 3.12 Hasil Perhitungan Probabilitas <i>Unigram</i> Dataset Bahasa Spanyol	29
Tabel 3.13 Hasil Perhitungan Probabilitas <i>Bigram</i> Dataset Bahasa Spanyol	31
Tabel 3.14 Matriks Perhitungan <i>Levenshtein Distance</i>	31
Tabel 3.15 Matriks Perhitungan <i>Levenshtein Distance</i>	39
Tabel 3.16 Tabel profile	39
Tabel 3.17 Tabel message	40
Tabel 3.18 Tabel otp	40

Tabel 4.1 Hasil Pengujian Bahasa Indonesia.....	
Tabel 4.2 Hasil Pengujian Bahasa Inggris.....	59
Tabel 4.3 Hasil Pengujian Bahasa Spanish.....	60
Tabel 4.4 Hasil Pengujian Bahasa Indonesia.....	62
Tabel 4.5 Hasil Pengujian Bahasa Inggris.....	63
Tabel 4.6 Hasil Pengujian Bahasa Spanish.....	64
Tabel 4.7 Hasil Pengujian Bahasa Indonesia.....	65
Tabel 4.8 Hasil Pengujian Bahasa Inggris.....	66
Tabel 4.9 Hasil Pengujian Bahasa Spanish.....	67
Tabel 4.10 Hasil Pengujian Bahasa Indonesia.....	69
Tabel 4.11 Hasil Pengujian Bahasa Inggris.....	70
Tabel 4.12 Hasil Pengujian Bahasa Spanish.....	71
Tabel 4.13 Hasil Pengujian Bahasa Indonesia.....	72
Tabel 4.14 Hasil Pengujian Bahasa Inggris.....	73
Tabel 4.15 Hasil Pengujian Bahasa Spanish.....	74
Tabel 4.16 Hasil Pengujian Bahasa Indonesia.....	75
Tabel 4.17 Hasil Pengujian Bahasa Inggris.....	76
Tabel 4.18 Hasil Pengujian Bahasa Spanish.....	77
Tabel 4.19 Akurasi Autocorrect Bahasa Indonesia.....	79
Tabel 4.20 Akurasi Autocorrect Bahasa Inggris.....	79
Tabel 4.21 Akurasi Autocorrect Bahasa Spanish.....	80

DAFTAR PERSAMAAN

Persamaan 2.1 <i>Levenshtein Distance</i>	8
Persamaan 2.2 <i>N-Gram Unigram</i>	9
Persamaan 2.3 <i>N-Gram Bigram</i>	9
Persamaan 2.4 <i>MLE Unigram</i>	10
Persamaan 2.5 <i>MLE Bigram</i>	10
Persamaan 2.6 <i>Akurasi</i>	11

ABSTRAK

Aplikasi chat kerap menghadapi tantangan dalam menjaga akurasi pesan akibat kesalahan pengetikan (*typo*) yang dilakukan oleh pengguna. Kesalahan ini dapat menyebabkan gangguan komunikasi, terutama ketika kata yang salah tidak diperbaiki dan diproses apa adanya. Penelitian ini bertujuan untuk mengimplementasikan sistem *autocorrect* menggunakan metode *N-Gram* (*Unigram* dan *Bigram*) serta *Levenshtein Distance* pada aplikasi chat yang terintegrasi dengan fitur penerjemahan otomatis. Metode *N-Gram* digunakan untuk menganalisis probabilitas kemunculan kata berdasarkan korpus data dari buku *Game of Thrones*, sementara *Levenshtein Distance* dimanfaatkan untuk menghitung jarak edit dan menentukan kandidat koreksi terdekat. Sistem ini mendukung tiga bahasa: Indonesia, Inggris, dan Spanyol, serta menggunakan *Google Translate API* untuk proses penerjemahan. Pengujian dilakukan melalui enam skenario berbeda, termasuk kesalahan ringan, berat, penggunaan simbol atau angka, serta validasi terhadap kata-kata kontekstual seperti nama dan lokasi. Hasil menunjukkan tingkat akurasi koreksi rata-rata sebesar 84% untuk Bahasa Indonesia, 89% untuk Bahasa Inggris, dan 92% untuk Bahasa Spanyol. Sistem juga mampu mempertahankan konteks kalimat serta menghindari koreksi yang tidak diperlukan terhadap kata valid seperti nama atau alamat. Kombinasi metode *N-Gram* dan *Levenshtein Distance* terbukti efektif dalam meningkatkan akurasi koreksi otomatis.

Kata Kunci: Autocorrect, N-Gram, Levenshtein Distance, Aplikasi Chat

ABSTRACT

Chat applications often face challenges in maintaining message accuracy due to user typing errors (typos). These mistakes can disrupt communication, especially when incorrect words are processed without prior correction. This study aims to implement an autocorrect system using the N-Gram method (Unigram and Bigram) and Levenshtein Distance in a chat application integrated with an automatic translation feature. The N-Gram method is used to analyze word occurrence probabilities based on a corpus from the Game of Thrones book, while Levenshtein Distance is utilized to calculate edit distances and identify the closest correction candidates. The system supports three languages: Indonesian, English, and Spanish, and leverages the Google Translate API for translation. Testing was conducted across six different scenarios, including minor and major errors, symbol or number usage, and validation of contextual words such as names and locations. The results show an average correction accuracy of 84% for Indonesian, 89% for English, and 92% for Spanish. The system also successfully preserves sentence context and avoids unnecessary corrections to valid words such as names or addresses. The combination of N-Gram and Levenshtein Distance proves effective in improving automatic correction accuracy.

Keywords: *Autocorrect, N-Gram, Levenshtein Distance, Chat Appl*