

**Sistem Koreksi Kesalahan Tulisan pada Teks Karangan Siswa
Sekolah Dasar Menggunakan Metode *N-Gram* dan
*Levenshtein Distance***

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh :

Muhammad Adi Murdilan

NIM : 09021281520138

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN TUGAS AKHIR

Sistem Koreksi Kesalahan Tulisan pada Teks Karangan Siswa Sekolah Dasar Menggunakan Metode *N-Gram* dan *Levenshtein Distance*

Oleh :

Muhammad Adi Murdilan

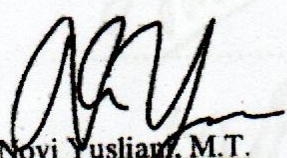
NIM : 09021281520138

Palembang, Desember 2020

Pembimbing I,

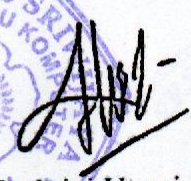

Dr. Abdiansah, S.Kom., M.Cs
NIP. 198410012009121005

Pembimbing II,


Novi Yusliany, M.T.
NIP. 198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika




Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

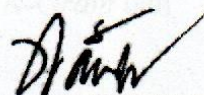
TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari selasa tanggal 29 Desember 2020 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

N a m a : Muhammad Adi Murdilan
N I M : 09021281520138
Judul : Sistem Koreksi Kesalahan Tulisan pada Teks Karangan Siswa Sekolah Dasar Menggunakan Metode *N-Gram* dan *Levenshtein Distance*.

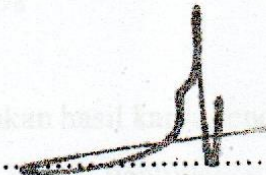
1. Ketua Penguji

Syamsuryadi, S.Si., M.Kom., Ph.D.
NIP. 197102041997021003



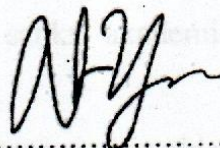
2. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs
NIP. 198410012009121005




3. Pembimbing II

Novi Yusliani, M.T.
NIP. 198211082012122001



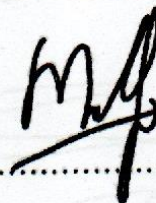
4. Penguji I

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



5. Penguji II

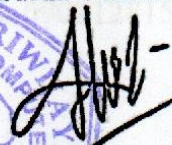
M Qurhanul Rizqie, S.Kom, M.T.
NIP. 1671060312870008



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Muhammad Adi Murdilan
NIM : 09021281520138
Program Studi : Teknik Informatika Reguler
Judul Skripsi : Sistem Koreksi Kesalahan Tulisan pada
Teks Karangan Siswa Sekolah Dasar
Menggunakan Metode *N-Gram* dan
Levenshtein Distance

Hasil Pengecekan Software *iThenticate /Turnitin* :

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Desember 2020

Muhammad Adi Murdilan
NIM. 09021281520138

MOTTO DAN PERSEMBAHAN

“Sesungguhnya sesudah kesulitan itu ada kemudahan.”
(QS. Al-Insyirah, 94:6)

“If I Die Trying, at Least I Tried”
(Monkey D. Luffy)

“Sesuatu yang **Tak Bernilai**, bisa menjadi **Tak Ternilai**”
(Murdilan)

Kupersembahkan Karya Tulis ini kepada :

- Orang Tuaku
- Adik-adikku
- Dosen Pembimbing
- Sahabat-sahabat
- Fakultas Ilmu Komputer, Unsri

**WRITING ERROR CORRECTION SYSTEM IN BASIC SCHOOL
STUDENTS 'TEXT USING *N-GRAM* AND *LEVENSHTEIN DISTANCE*
METHODS**

ABSTRACT

One of the crucial things in making documents and texts is a typographical error. There are several types of typographical errors, namely insertion errors, deletion errors and substitution errors. In the text written by elementary school students, typographical errors occur because students do not understand how to make essays in Indonesian which are standardized and in accordance with EYD rules. Therefore, we need a system to detect word errors. In this study, checking was carried out using the Dictionary Lookup method and providing word candidates using the Levenshtein Distance algorithm. The word candidates will then determine the ranking by breaking the words using the N-Gram method into Bigram form and calculating the similarity value using the Cosine Similarity algorithm. The test results of this system obtained different MMR values for each type of error. For the insertion type, the MMR value is 88,89%; then the type of substitution obtained a the MMR value is 46,96%; and for the deletion type, the MMR value is 93,33%.

Keywords : N-gram, Levenshtein Distance, MMR, Spelling Correction

**SISTEM KOREKSI KESALAHAN TULISAN PADA TEKS KARANGAN
SISWA SEKOLAH DASAR MENGGUNAKAN METODE N-GRAM DAN
LEVENSHTEIN DISTANCE**

ABSTRAK

Salah satu hal krusial dalam pembuatan dokumen maupun teks adalah kesalahan penulisan kata atau yang disebut *typographical error*. Terdapat beberapa jenis *typographical error* yaitu *insertion error*, *deletion error* serta *substitution error*. Pada teks karangan siswa sekolah dasar, *typographical error* terjadi karena belum pahamnya siswa dalam pembuatan karangan dengan bahasa Indonesia yang baku dan sesuai dengan kaidah EYD. Oleh karena itu, diperlukan suatu sistem untuk mendeteksi kesalahan penulisan kata. Pada penelitian ini, pengecekan dilakukan dengan menggunakan metode *Dictionary Lookup* dan memberikan kandidat kata menggunakan algoritma *Levenshtein Distance*. Kandidat kata kemudian akan dilakukan penentuan ranking dengan memecah kata dengan metode N-Gram menjadi bentuk *Bigram* dan menghitung nilai kemiripan dengan menggunakan algoritma *Cosine Similarity*. Hasil Pengujian dari sistem ini didapatkan MMR yang berbeda-beda untuk setiap jenis kesalahan. Untuk jenis *insertion* didapat nilai MMR sebesar 88,89%; selanjutnya jenis *substitution* didapat nilai MMR sebesar 46,96%; serta untuk jenis *deletion* didapat nilai MMR sebesar 93,33%.

Kata kunci : N-gram, Levenshtein Distance, MMR, Spelling Correction

KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala rahmat dan karunia-Nya penulis dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika pada Fakultas Ilmu Komputer, Universitas Sriwijaya.

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan bantuan dan dukungan baik dalam materil maupun moril selama proses pembuatan tugas akhir ini. Adapun penulis merujuk secara khusus pihak yang telah membantu sebagai berikut :

1. Bapak Aswan dan ibu Eftiah selaku orang tua penulis, Aziz dan Sastri selaku saudara penulis, serta keluarga besar yang telah memberikan dukunan baik moril maupun materil kepada penulis
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Sriwijaya
3. Ibu Avi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya sekaligus penguji I Tugas Akhir
4. Bapak Danny Matthew Saputra, S.T., M.Sc. selaku pembimbing akademik
5. Bapak Dr. Abdiansah., S.Kom., M.Cs selaku pembimbing I Tugas Akhir dan ibu Novi Yusliani, M.T. selaku pembimbing II Tugas Akhir yang telah memberikan pengarahan, bimbingan, bantuan serta masukan kepada penulis sehingga Tugas Akhir ini dapat diselesaikan dengan baik.
6. Bapak M. Qurhanul Rizkqie, S.Kom,M.T. selaku penguji II Tugas Akhir yang telah banyak memberikan saran dan kritik serta arahan selama penulisan Tugas Akhir
7. Seluruh Ibu dan Bapak Dosen Jurusan Teknik Informatika Fakultas Komputer Universitas Sriwijaya yang telah memberikan ilmu pengetahuan kepada penulis selama masa perkuliahan

8. Seluruh karyawan dan karyawan Fakultas Ilmu Komputer Universitas Sriwijaya yang telah membantu penulis dalam urusan administrasi selama kegiatan akademik
9. Seluruh sahabat dan teman-teman yang telah memberikan support sehingga penulis dapat menyelesaikan Tugas Akhir, terkhusus Icad, Adnan, Faruq, Acong, Kamsan, Derry, Ikhsan, Tomi
10. UKM Unsri Mengajar yang telah memberikan banyak pengalaman, wawasan, serta hal-hal lain yang penulis rasakan sebagai kebaikan dan kebermanfaatan
11. Ibu Sayang Ajeng Mardhiyah, S.Psi., M.Si. yang telah menjadi dosen “pembimbing kehidupan” penulis selama pengerjaan Tugas Akhir
12. Sayekti Dyah Iswari, S.Kom. yang telah menjadi sahabat penulis sejak hari pertama PK2 dan selalu menjadi motivator untuk penulis sehingga penulis bisa menjadi sosok yang kuat dan selalu berusaha memperbaiki diri menjadi lebih baik

Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya maupun pembaca pada umumnya. Serta dapat menjadi referensi dan rujukan bagi hal-hal yang bermanfaat. Penulis menyadari bahwa dalam proses penyelesaian tugas akhir ini, terdapat beberapa penjelasan yang kurang sempurna. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun agar tugas akhir ini menjadi karya tulis yang sempurna supaya terciptanya bekal pengetahuan yang baik bagi peneliti di masa depan.

Palembang, 29 Desember 2020

Muhammad Adi Murdilan

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
HALAMAN PENYATAAN	iiv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	v
BAB 1 PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang.....	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian.....	I-3
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-4
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Kajian Pustaka	II-1
2.3 Teks	II-3
2.4 Teks Mining.....	II-4
2.4.1 Pengertian Teks <i>Mining</i>	II-4
2.4.2 <i>Information Retrieval</i>	II-6
2.4.3 <i>Natural Language Processing</i>	II-8
2.4.4 Teks <i>Preprocessing</i>	II-8
2.5 <i>Spelling checker</i>	II-10
2.6 Metode N-gram	II-10
2.7 <i>Term Frequency – Inverse Document Frequency (tf-idf)</i>	II-11
2.8 <i>Consine Similarity</i>	II-11

2.9	Metode <i>Levensthein</i>	II-14
2.10	<i>Mean Reciprocal Rank</i> (MRR).....	II-16
BAB III METODE PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data.....	III-2
3.3	Analisa Kebutuhan dan Perancangan Sistem	III-3
3.3.1	Analisa Kebutuhan	III-3
3.3.2	Perancangan Sistem.....	III-4
3.4	Implementasi	III-4
3.5	Pengujian dan Analisis	III-5
3.6	Kesimpulan.....	III-6
3.7	Manajemen Proyek Penelitian.....	III-6
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis.....	IV-1
4.2.2	Kebutuhan Sistem	IV-2
4.2.3	Analisi dan Desain	IV-3
4.2.3.1	Analisis Kebutuhan Perangkat Lunak	IV-4
4.2.3.2	Analisis Data	IV-5
4.2.3.3	Analisis <i>Preprocessing</i>	IV-6
4.2.3.4	Analisis Identifikasi Kesalahan <i>Typographical Error</i>	IV-10
4.2.3.5	Analisis <i>Levenshtein Distance</i>	IV-12
4.2.3.6	Analisis <i>N-gram</i>	IV-19
4.2.3.7	Analisis <i>Mean Reciprocal Rank</i> (MMR).....	IV-24
4.2.3.8	Desain Perangkat Lunak.....	IV-25
4.3	Fase Elaborasi.....	IV-30
4.3.1	Pemodelan Bisnis	IV-30
4.3.2	Perancangan Data.....	IV-30
4.3.3	Perancangan Antarmuka	IV-31
4.3.4	Kebutuhan Sistem	IV-31
4.3.5	Diagram Aktivitas (<i>Activity Diagram</i>)	IV-32
4.4	Fase Konstruksi	IV-35
4.4.1	Kebutuhan Sistem	IV-35

4.5	Fase Transisi	IV-35
4.5.1	Pemodelan Bisnis	IV-35
4.5.2	Kebutuhan Sistem	IV-36
4.5.3	Rencana Pengujian	IV-36
4.5.3.1	Rencana Pengujian <i>Use case</i> Melakukan Input data.....	IV-36
4.5.3.2	Rencana Pengujian <i>Use case</i> Melakukan Proses dan Ouput Data	IV-37
4.5.4	Implementasi	IV-38
4.5.4.1	Pengujian <i>Use case</i> Melakukan Input data.....	IV-38
4.5.4.2	Pengujian <i>Use case</i> Melakukan Proses dan Output Data....	IV-39
4.5	Kesimpulan.....	IV-40
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Data Hasil Konfigurasi.....	V-1
5.3	Analisis Hasil Penelitian.....	V-7
5.4	Kesimpulan.....	V-9
BAB VI KESIMPULAN DAN SARAN		VI-1
6.1	Kesimpulan.....	V1-1
6.2	Saran.....	V1-2
DAFTAR PUSTAKA		

DAFTAR TABEL

	Halaman
Tabel II-1. Kajian Pustaka.....	II-2
Tabel IV-1. Karangan Siswa Sebagai Data Input Sistem	IV-2
Tabel IV-2. Kebutuhan Fungsional dan Non-Fungsional	IV-3
Tabel IV-3. Hasil Proses <i>Case Folding</i>	IV-8
Tabel IV-4. Hasil Proses <i>Remove Punctuation</i>	IV-9
Tabel IV-5. Hasil Proses <i>Tokenization</i>	IV-9
Tabel IV-6. Perhitungan Nilai Edit Distance Kata Typographical Error Terhadap Kandidat Kata Pertama	IV-13
Tabel IV-7. Matriks Hasil Perhitungan <i>Edit Distance</i> Kandidat Pertama	IV-14
Tabel IV-8. Perhitungan Nilai Edit Distance Kata Typographical Error Terhadap Kandidat Kata Kedua	IV-15
Tabel IV-9. Matriks Hasil Perhitungan <i>Edit Distance</i> Kandidat Kedua.....	IV-16
Tabel IV-10. Perhitungan Nilai Edit Distance Kata Typographical Error Terhadap Kandidat Kata Ketiga	IV-17
Tabel IV-11. Matriks Hasil Perhitungan <i>Edit Distance</i> Kandidat Ketiga.....	IV-18
Tabel IV-12. Perhitungan Nilai <i>tf</i> pada Kata Typographical Error dan Kandidat Kata.....	IV-20
Tabel IV-13. Perhitungan Nilai <i>idf</i> pada Kata Typographical Error dan Kandidat Kata.....	IV-21
Tabel IV-14. Perhitungan Nilai <i>tf-idf</i> pada Kata Typographical Error dan Kandidat Kata.....	IV-21
Tabel IV-15. Perhitungan <i>Cosine Similarity</i> pada Kata <i>Typographical Error</i> dan Kandidat Kata.....	IV-22
Tabel IV-16. Hasil <i>Cosine Similarity</i> pada Semua Kandidat Kata	IV-22
Tabel IV-17. Contoh Perhitungan MMR	IV-24
Tabel IV-18. Definisi Aktor	IV-26
Tabel IV-19. Definisi <i>Use Case</i>	IV-27
Tabel IV-20. Skenario <i>Use Case Input Data</i>	IV-27
Tabel IV-21. Skenario <i>Use Case Proses</i>	IV-28
Tabel IV-22. Skenario <i>Output Data</i>	IV-29
Tabel IV-23. Rancangan Data.....	IV-31
Tabel IV-24. Rencana Pengujian <i>Use Case</i> Melakukan <i>Input Data</i>	IV-37
Tabel IV-25. Rencana Pengujian <i>Use Case</i> Melakukan Proses dan <i>Output Data</i>	IV-37
Tabel IV-26. Pengujian <i>Use Case</i> Melakukan <i>Input Data</i>	IV-38
Tabel IV-27. Pengujian <i>Use Case</i> Melakukan Proses dan <i>Output Data</i>	IV-39
Tabel V-1. Hasil Perhitungan Sistem.....	V-2
Tabel V-2. Frekuensi Hasil	V-7
Tabel V-3. Hasil Perhitungan Nilai MMR.....	V-8

DAFTAR GAMBAR

	Halaman
Gambar II-1. Proses pada Teks Mining	II-5
Gambar II-2. Bagian-bagian <i>Information System Retrieval (IR)</i>	II-6
Gambar II-3. Tahap Preprocessing	II-9
Gambar II-4 .Derajat <i>Cosine Similarity</i>	II-13
Gambar III-1. Blok Diagram Metodologi Penelitian	III-1
Gambar III-2. Diagram Gantt Penelitian	III-6
Gambar IV-1. Alur Identifikasi dan Koreksi <i>Typographical Error</i>	IV-4
Gambar IV-2. Alur Tahapan <i>Text Preprocessing</i>	IV-7
Gambar IV-3. Alur Proses Identifikasi <i>Typographical Error</i>	IV-11
Gambar IV-4. Diagram <i>Use Case</i>	IV-26
Gambar IV-5. Rancangan Antarmuka Perangkat Lunak	IV-31
Gambar IV-6. Diagram Aktivitas <i>Input Data</i>	IV-33
Gambar IV-7. Diagram Aktivitas Proses dan <i>Output Data</i>	IV-34

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini berisi pokok-pokok pikiran yang melandasi pembuatan skripsi. Pokok-pokok pikiran tersebut meliputi latar belakang, rumusan masalah, tujuan, dan manfaat yang diperoleh dalam melakukan penelitian serta batasan masalah.

1.2 Latar Belakang

Salah satu hal krusial dalam pembuatan dokumen teks adalah kesalahan penulisan kata atau yang disebut *typographical error*. Terdapat beberapa jenis *typographical error* yaitu *insertion error*, *deletion error* serta *substitution error*. Dalam pembuatan sebuah karangan siswa menghadapi beberapa kesulitan yang diantaranya adalah: (a) menyusun judul dengan baik; (b) menentukan ide pokok cerita; (c) mengembangkan paragraf meliputi kesulitan siswa dalam menyusun kalimat dengan struktur kalimat yang tepat, serta pemilihan kata/ diksi yang sesuai; (d) penggunaan ejaan dan tanda baca meliputi penulisan kata berimbuhan dan penggunaan huruf kapital (Khotimah & Suryandari, 2016). Dalam mendeteksi *typographical error* pada teks dibutuhkan suatu aplikasi disebut *spelling checker*. Kebutuhan akan *spelling checker* dalam dokumen berbahasa Indonesia sangat tinggi, namun ketersediaan aplikasi yang berguna dalam mendeteksi *typographical error* teks masih sedikit terutama untuk mengecek kesalahan pada karangan siswa sekolah dasar.

Beberapa metode pernah digunakan untuk membuat sistem identifikasi *typographical error* pada teks seperti *N-gram* (Wardhana et al., 2011), *Hidden*

Markov Model (HMM) (Soleh & Purwarianti, 2011), *Forward-Reversed Dictionary* (Aqsath Rasyid et al., 2011), *Morphologically Analyzer* (Soleh & Purwarianti, 2011), *Bayes* (Setiadi, 2013), *Maximum Likelihood Estimation* (Christanti Mawardi et al., 2018), *Minimum Edit Distance* (Safri et al., 2011), *Similarity Key* (Martins & Silva, 2004), dan *Neural Net* (Augusfian et al., 2019).

Penelitian untuk pembuatan *spelling correction tool* menggunakan metode *N-gram* dengan beberapa macam pendekatan menghasilkan akurasi yang bagus. Hasil yang lebih unggul didapatkan dari nilai N sebesar 2, yang disebut dengan *bigram*, daripada dengan penggunaan N sebesar 3 yang disebut dengan *trigram* (Fahma, 2018). Penggunaan data uji berupa dataset bahasa Inggris menghasilkan akurasi sebesar 84% untuk *bigram* dan 73% untuk *trigram* (Ahmed, Luca dan Nürnbergger, 2009). Untuk pengoreksian ejaan kata dalam teks bahasa Indonesia, pengoreksian ejaan bisa memberikan hasil yang baik saat digunakan dengan bantuan dari *user* untuk memilih kata saran yang ditunjang dengan waktu eksekusi yang masih cukup cepat, namun masih kurang sempurna pengoreksiannya jika seluruhnya dilakukan otomatis oleh sistem (Wardhana et al., 2011).

Selain itu, terdapat penelitian menggunakan metode *Levenshtein Distance* dengan pendekatan yang lebih baik yaitu penambahan metode *Dictionary Lookup* (Haldar & Mukhopadhyay, 2011). Metode *Levenshtein Distance* akan mengurangi jumlah kata yang tidak dikenali sebanyak 66 kata dan metode yang dimodifikasi yaitu metode *Levenshtein Distance* mengurangi jumlah kata yang tidak dikenali dari 93 menjadi 52 sehingga baik untuk sistem identifikasi kesalahan tulisan.

Pada penelitian ini metode *N-gram* dan metode *Levenshtein Distance* akan digunakan dalam pembuatan sebuah sistem koreksi kesalahan tulisan pada teks karangan siswa sekolah dasar. Selanjutnya penelitian ini dikembangkan dengan penggunaan metode *N-gram* yang digunakan untuk menghitung nilai *cosine similarity* dalam menentukan *ranking* kandidat kata sebagai hasil keluaran dari sistem. Lalu Metode *Levenshtein Distance* digunakan untuk menentukan kandidat kata untuk setiap kesalahan tulisan yang teridentifikasi oleh sistem. Diharapkan sistem ini dapat membantu memberikan gambaran yang cukup baik terhadap koreksi kesalahan tulisan pada teks karangan berdasarkan hasil pada sistem tersebut.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang ada, maka rumusan masalah yang dapat dibuat adalah sebagai berikut:

1. Bagaimana melakukan indentifikasi kesalahan tulisan pada teks karangan siswa sekolah dasar menggunakan LD dan N-gram.
2. Berapa nilai *Mean Reciprocal Rank* (MMR) dari setiap jenis typographical error pada indentifikasi kesalahan tulisan menggunakan LD dan N-gram.

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan pencapaian di antaranya adalah:

1. Membuat perangkat lunak untuk mendeteksi kesalahan tulisan pada teks karangan siswa sekolah dasar menggunakan LD dan N-gram.
2. Mengetahui tingkat nilai *Mean Reciprocal Rank* (MMR) dari setiap jenis typographical error pada indentifikasi kesalahan tulisan teks.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Bagi Masyarakat
 - Memberikan sarana untuk mengidentifikasi kesalahan tulisan pada teks karangan siswa sekolah dasar.
2. Bagi Peneliti Lain
 - Dapat digunakan untuk studi penelitian selanjutnya terkait identifikasi kesalahan tulisan pada teks lainnya.
3. Bagi Institusi
 - Penelitian ini diharapkan dapat memberikan kontribusi penambahan ilmu pengetahuan, serta menjadi bahan bacaan untuk referensi mahasiswa lain.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Metode yang digunakan untuk identifikasi kesalahan tulisan adalah *Dictionary Lookup*.
2. Metode N-gram yang digunakan berbasis nilai $N=2$ dan perhitungan *cosine similarity*.
3. Data uji sistem yang digunakan adalah 30 teks karangan siswa kelas 3 sampai dengan kelas 6 sekolah dasar.

1.7 Sistematika Penulisan

Penyusunan skripsi ini disusun dengan sistematika penulisan sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai pokok-pokok pikiran yang melandasi pembuatan penelitian, seperti latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini dijelaskan mengenai beberapa teori terkait teks, metode *N-gram*, dan metode *Levenshtein Distance*. Penelitian-penelitian sebelumnya menyangkut implementasi dari sistem juga akan dibahas dalam kajian pustaka.

BAB III. METODOLOGI PENELITIAN

Pada bab ini membahas mengenai tahapan yang akan dilaksanakan pada penelitian. Seperti pengumpulan data, analisis data, serta perancangan sistem yang akan dibangun. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci mengacu pada suatu kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini membahas mengenai analisis dan perancangan perangkat lunak yang akan digunakan sebagai alat penelitian. Dimulai dari pengumpulan dan analisa kebutuhan, rancangan dan konstruksi perangkat lunak serta pengujian untuk memastikan semua kebutuhan pengembangan perangkat lunak sesuai dengan kebutuhan.

BAB V. HASIL DAN ANALISA PENELITIAN

Pada bab ini diuraikan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Tabel hasil pengujian serta analisisnya disajikan sebagai basis dari kesimpulan yang akan diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga saran yang diberikan berdasarkan hasil dari penelitian.

1.8 Kesimpulan

Dari pendahuluan ini, telah diuraikan secara umum pokok-pokok pikiran yang melandasi penelitian yang akan dilakukan, meliputi latar belakang dan rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

DAFTAR PUSTAKA

- Aqsath Rasyid, N., Kamayani, M., Reinanda, R., Simbolon, S., Soleh, M. Y., & Purwarianti, A. (2011). Application of document spelling checker for Bahasa Indonesia. *ICACSYS 2011 - 2011 International Conference on Advanced Computer Science and Information Systems, Proceedings, January*, 249–252.
- Ariyani, N. H., Sutardi, & Ramadhan, R. (2016). Aplikasi Pendeteksi Kemiripan Isi Teks Dokumen Menggunakan Metode Levenshtein Distance. *SemanTIK, Vol 2(1)*, 279–286.
<http://ojs.uho.ac.id/index.php/semantik/article/download/1030/661>
- Augusfian, F., M, V. C., Hendryli, J., & Naga, D. S. (2019). *Indonesia Dengan Damerau Levenshtein. 2*, 144–152.
- Bunyamin, H. (2005). *Information Retrieval System dengan Metode Latent Semantic Indexing*.
- Bunyamin, H., Negara, C. P., Informasi, F. T., & Maranatha, U. K. (2008). Aplikasi Information Retrieval (IR) CATA Dengan Metode Generalized Vector Space Model. *Jurnal Informatika, 4(1)*, 29–38.
- Christanti Mawardi, V., Susanto, N., & Santun Naga, D. (2018). Spelling Correction for Text Documents in Bahasa Indonesia Using Finite State Automata and Levenshtein Distance Method. *MATEC Web of Conferences, 164*. <https://doi.org/10.1051/mateconf/201816401047>
- Fahma, A. I. (2018). Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 2(1)*, 53–62.
- Gomma, W. H., & Fahmy, A. A. (2013). A Survey of Text Similarity Approaches. *International Journal of Computer Applications, 68(13)*, 13–18.
- Haldar, R., & Mukhopadhyay, D. (2011). *Levenshtein Distance Technique in Dictionary Lookup Methods: An Improved Approach. June*.
<http://arxiv.org/abs/1101.1232>
- Khotimah, H., & Suryandari, K. C. (2016). Analisis Kesulitan Menulis Karangan

- Pada Siswa Kelas Iv Sdn 2 Panjer. *Seminar Nasional Inovasi Pendidikan*, 491–500.
- Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016*. <https://doi.org/10.1109/CITSM.2016.7577578>
- Martins, B., & Silva, M. J. (2004). Spelling correction for search engine queries. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3230(January), 372–383. https://doi.org/10.1007/978-3-540-30228-5_33
- Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *Journal of Machine Learning Technology*, 2(1), 37–63.
- Pratama, B. P., & Pamungkas, S. A. (2016). Analisis Kinerja Algoritma Levenshtein Distance dalam Mendeteksi Kemiripan Dokumen Teks. *Jurnal "LOG!K@,"* 6(2), 131–143.
- Putera Utama Siahaan, A., Aryza, S., Hariyanto, E., . R., Hasudungan Lubis, A., Ikhwan, A., & Len Eh Kan, P. (2018). Combination of levenshtein distance and rabin-karp to improve the accuracy of document equivalence level. *International Journal of Engineering & Technology*, 7(2.27), 17. <https://doi.org/10.14419/ijet.v7i2.27.12084>
- Safri, M. F., Dayawati, R. N., & Wibowo, A. T. (2011). *Implementasi dan analisis algoritma edit distance untuk spelling checker pada document teks bahasa indonesia*. 1–5.
- Satya, K. P. N. V., & Murthy, J. V. R. (2012). Clustering based on cosine similarity measure. *International Journal of Engineering Science & Advanced Technology*, 3, 508–512.
- Setiadi, I. (2013). Damerau-Levenshtein Algorithm and Bayes Theorem for Spell Checker Optimization. *Bandung Institute of Technology, November*, 1–6. <https://doi.org/10.13140/2.1.2706.4008>
- Singh, A., & Sharma, M. (2016). Spell checking and error correcting system for

text paragraphs written in Punjabi or Hindi language using hybrid approach. *International Journal of Engineering Sciences and Research Technology*, 5(4), 399–408. <https://www.ijarset.com/upload/2015/november/9-IJARSET-Amanjotkaur.pdf>

Soepomo, P. (2014). Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes. *Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes*, 2(3), 73–83. <https://doi.org/10.12928/jstie.v2i3.2877>

Soleh, M. Y., & Purwarianti, A. (2011). A non word error spell checker for Indonesian using morphologically analyzer and HMM. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011, July 2019*. <https://doi.org/10.1109/ICEEI.2011.6021514>

Wangsanegara, N. K., & Subaeki, B. (2015). Implementasi Natural Language Processing Dalam Pengukuran Ketepatan Ejaan Yang Disempurnakan (Eyd) Pada Abstrak Skripsi Menggunakan Algoritma Fuzzy Logic. *Jurnal Teknik Informatika*, 8(2). <https://doi.org/10.15408/jti.v8i2.3185>

Wardhana, W. S., Wirayuda, T. A. B., & Shaufiah. (2011). *Pengoreksian Ejaan Kata Menggunakan Metode N-Gram (Studi Kasus Dokumen Teks Berbahasa Indonesia)*. 0–6.

Yulianto, M. M., Arifudin, R., & Alamsyah, A. (2018). Autocomplete and Spell Checking Levenshtein Distance Algorithm To Getting Text Suggest Error Data Searching In Library. *Scientific Journal of Informatics*, 5(1), 75. <https://doi.org/10.15294/sji.v5i1.14148>

Zhang, W. (2015). *Comparing the Effect of Smoothing and N-gram Order : Finding the Best Way to Combine the Smoothing and Order of N-gram*. 1–33.