

**Analisa Perbandingan Algoritma *Damerau-Levenshtein Distance*
dan *Soundex Similarity* Pada Pengoreksian Ejaan Kata Otomatis**

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



Oleh:

Abson Hadi
09121002015

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2019**

LEMBAR PENGESAHAN TUGAS AKHIR

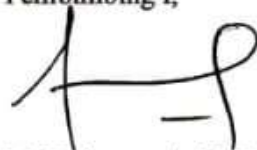
ANALISA PERBANDINGAN *ALGORITMA DAMERAU-LEVENSHTEIN DISTANCE* DAN *SOUNDEX SIMILARITY* PADA PENGOREKSIAN EJAAN KATA OTOMATIS

Oleh :

ABSON HADI
NIM : 09121002015

Indralaya, Juli 2019

Pembimbing I,



M. Fachrurrozi, M. T.
NIP. 198005222008121002

Pembimbing II,




Novi Yoslina, M. T.
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika,




Rifkie Primartha, M. T.
NIP. 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Rabu tanggal 31 Juli 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Abson Hadi

NIM : 09121002015

Judul : Analisa Perbandingan Algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity* Pada Pengoreksian Ejaan Kata Otomatis

1. Pembimbing I

M. Fachrurrozi, M.T.

NIP. 198005222008121002

2. Pembimbing II

Novi Yusliani, M.T.

NIP. 198211082012122001

3. Penguji I

Ali Ibrahim, M. T.

NIP. 198407212010031201

4. Penguji II

M. Naufal Rachmatullah, M. T.

NIP.

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Pramadha, M. T.

NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Abson Hadi
NIM : 09121002015
Program Studi : Teknik Informatika
Judul Skripsi : Analisa Perbandingan Algoritma *Damerau-Levenshtein Distance*
dan *Soundex Similarity* Pada Pengoreksian Ejaan Kata Otomatis
Hasil Pengecekan Software *iThenticate Turnitin* : 14 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek 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, Juli 2019



Abson Hadi
NIM. 09121002015

Motto :

- *Just do it even if it's hard, everything will pass soon*
- *You don't know what will happen in the future, prepare your best*

Kupersembahkan karya tulis ini kepada :

- *Kedua orang tua dan adikku*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

COMPARATIVE ANALYSIS OF DAMERAU-LEVENSHTEIN DISTANCE AND SOUNDEX SIMILARITY ALGORITHM IN AUTOMATIC SPELLING CORRECTION

By:
Abson Hadi
09021181520023

ABSTRACT

Damerau-Levenshtein Distance and Soundex Similarity are two methods that can be used to correct word spelling errors automatically. The difference between the two algorithms lies in the basic correction where the Damerau-Levenshtein Distance algorithm corrects word spelling errors based on word match, while Soundex Similarity algorithm corrects word spelling errors based on sound similarity. Although it is known that the two algorithms have differences, the most effective algorithm in correcting spelling errors of words is still unknown from the two algorithms. Because of this, a study was conducted to analyze the comparison of the two algorithms in automatic word spelling correction. The testing method is done by giving 50 words with spelling errors to be corrected automatically using the Damerau-Levenshtein Distance and Soundex Similarity algorithm. The results showed the Damerau-Levenshtein Distance method managed to get an accuracy of 72% with the success of word correction of 36 words. While Soundex Similarity algorithm gets an accuracy of 68% with the success of word correction of 34 words in automatic word spelling corrections.

Keywords : *Spelling Correction, Damerau-Levenshtein Distance, Soundex Similarity*

Pembimbing I,



M. Fachrurrozi, M. T.
NIP. 198005222008121002

Indralaya, Juli 2019
Pembimbing II,



Novi Yushani, M. T.
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika,



Rikie Primartha, M. T.
NIP. 197706012009121004

ANALISA PERBANDINGAN ALGORITMA *DAMERAU-LEVENSHTEIN DISTANCE* DAN *SOUNDEX SIMILARITY* PADA PENGOREKSIAN EJAAN KATA OTOMATIS

Oleh :
Abson Hadi
09121002015

ABSTRAK

Damerau-Levenshtein Distance dan *Soundex Similarity* merupakan dua metode yang dapat digunakan dalam mengoreksi kesalahan ejaan kata secara otomatis. Perbedaan kedua algoritma terletak pada dasar pengoreksian dimana algoritma *Damerau-Levenshtein Distance* mengoreksi kesalahan ejaan kata berdasarkan kecocokan kata, sedangkan algoritma *Soundex Similarity* mengoreksi kesalahan ejaan kata berdasarkan kemiripan bunyi. Walaupun telah diketahui bahwa kedua algoritma memiliki perbedaan, algoritma yang paling efektif dalam mengoreksi kesalahan ejaan kata masih belum diketahui dari kedua algoritma tersebut. Dari hal tersebut, dilakukan penelitian untuk menganalisa perbandingan dari kedua algoritma dalam koreksi ejaan kata otomatis. Metode pengujian dilakukan dengan memberikan 50 kata dengan kesalahan ejaan untuk dikoreksi secara otomatis menggunakan algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity*. Hasil penelitian menunjukkan metode *Damerau-Levenshtein Distance* berhasil mendapatkan akurasi sebesar 72% dengan keberhasilan koreksi kata sebanyak 36 kata. Sedangkan algoritma *Soundex Similarity* mendapatkan akurasi sebesar 68% dengan keberhasilan koreksi kata sebanyak 34 kata pada pengoreksian ejaan kata secara otomatis.

Kata Kunci : *Spelling Correction*, *Damerau-Levenshtein Distance*, *Soundex Similarity*

Pembimbing I,



M. Fachrurrozi, M. T.
NIP. 198005222008121002

Indrataya, Juli 2019
Pembimbing II,



Novi Yushiani, M. T.
NIP. 19821082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M. T.
NIP. 197706012009121004

KATA PENGANTAR

Bismillahirrahmanir rahtim

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, Rusman dan Mimi Helmi, adikku, Admen Ardio serta seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M. T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika dan Ibu Hardini Novianti, M. T. selaku Sekretaris Jurusan Teknik Informatika.
3. Bapak M. Fachrurrozi, M.T selaku dosen pembimbing I dan Ibu Novi Yusliani, M.T selaku pembimbing II, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Samsuryadi, M. Kom., Ph. D. selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak Ali Ibrahim, M. T. selaku dosen penguji I, dan Bapak Naufal Rahmatullah selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Johni, Niko, Aud, Genda, Putri, Dara, serta seluruh teman-teman jurusan Teknik Informatika yang telah saling berbagi selama masa perkuliahan ini.
9. BPH HMIF, Wifi, Digital Creative yang telah memberikan ruang bagi Penulis untuk berprestasi dan berkarya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Juli 2019

Abson Hadi

DAFTAR ISI

| | |
|---|-------|
| HALAMAN JUDUL..... | i |
| HALAMAN PENGESAHAN..... | ii |
| HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR..... | iii |
| HALAMAN PERNYATAAN | iv |
| MOTTO DAN PERSEMBAHAN | v |
| ABSTRACT..... | vi |
| ABSTRAK..... | vii |
| KATA PENGANTAR | viii |
| DAFTAR ISI | ix |
| DAFTAR TABEL..... | xii |
| DAFTAR GAMBAR | xiv |
| DAFTAR LAMPIRAN..... | xv |
| BAB I PENDAHULUAN | I-1 |
| 1.1 Pendahuluan | I-1 |
| 1.2 Latar Belakang Masalah..... | 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-5 |
| BAB II KAJIAN LITERATUR | II-1 |
| 2.1 Pendahuluan | II-1 |
| 2.2 Landasan Teori | II-1 |
| 2.2.1 Kesalahan Ejaan (<i>Spelling Error</i>) | II-1 |
| 2.2.2 <i>Spelling Correction</i> | II-2 |
| 2.2.3 <i>Approximate String Matching</i> | II-3 |
| 2.2.4 Praproses | II-3 |
| 2.2.5 <i>Algoritma Levenshtein Distance</i> | II-4 |
| 2.2.6 <i>Algoritma Damerau-Levenshtein Distance</i> | II-6 |
| 2.2.7 <i>Soundex Similarity</i> | II-9 |
| 2.2.8 Pengukuran Hasil | II-10 |
| 2.2.9 <i>Rational Unified Process (RUP)</i> | II-11 |

| | |
|---|--------------|
| 2.3 Penelitian Lain yang Relevan | II-13 |
| 2.3.1 Tusty Nadia Maghfira et all.2017.“Deteksi Kesalahan Ejaan dan Penentuan Rekomendasi Koreksi Kata Yang Tepat Pada Dokumen Jurnal JTIHK Menggunakan Dictionary Lookup dan Damerau-Levensthein Distance”.Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.Vol 1 | II-13 |
| 2.3.2 Ahmad Pahdi.2016.”Koreksi Ejaan Istilah Komputer Berbasis Kombinasi Algoritma Damerau-Levensthein dan Algoritma Soundex”. Journal Sentra Penelitian Engineering dan Edukasi.Vol 8..... | II-14 |
| 2.4 Kesimpulan..... | II-15 |
| BAB III METODOLOGI PENELITIAN..... | III-1 |
| 3.1 Pendahuluan | III-1 |
| 3.2 Data..... | III-1 |
| 3.2.1 Jenis Data dan Sumber Data..... | III-1 |
| 3.2.2 Metode Pengumpulan Data | III-1 |
| 3.3 Tahapan Penelitian | III-1 |
| 3.3.1 Kerangka Kerja..... | III-2 |
| 3.3.2 Kriteria Pengujian..... | III-4 |
| 3.3.3 Format Data Pengujian..... | III-4 |
| 3.3.4 Alat yang Digunakan dalam pengujian | III-5 |
| 3.3.5 Penelitian | III-6 |
| 3.3.6 Analisis Hasil Pengujian | III-6 |
| 3.4 Metode Pengembangan Perangkat Lunak | III-7 |
| 3.5 Penjadwalan Penelitian..... | III-15 |
| BAB IV PENGEMBANGAN PERANGKAT LUNAK..... | IV-1 |
| 4.1 Pendahuluan | IV-1 |
| 4.2 <i>Fase Inpesi</i> | IV-1 |
| 4.2.1 Pemodelan Bisnis | IV-1 |
| 4.2.2 Kebutuhan Sistem..... | IV-2 |
| 4.2.2.1 Fitur Koreksi | IV-2 |
| 4.2.3 Analisis dan Desain | IV-4 |
| 4.2.3.1 Analisis Data..... | IV-4 |
| 4.2.3.2 Analisis Dokumen Masukan | IV-5 |
| 4.2.3.3 Analisis <i>Pre-processing</i> | IV-6 |
| 4.2.3.4 Analisis <i>Casefolding</i> | IV-6 |
| 4.2.3.5 Analisis <i>Tokenizing</i> | IV-6 |
| 4.2.3.6 Analisis Pencocokan Kata..... | IV-7 |

| | |
|---|--------------|
| 4.2.3.7 Analisis Pengkodean <i>Soundex</i> | IV-8 |
| 4.2.3.8 Analisis Pembobotan Damerau-Levenshtein Distance | IV-9 |
| 4.2.3.9 Analisis Perangkat Lunak | IV-13 |
| 4.2.3.9.1 Model <i>Use Case</i> | IV-14 |
| 4.3 Fase <i>Elaborasi</i> | IV-18 |
| 4.3.1 Pemodelan Bisnis | IV-18 |
| 4.3.1.1 Perancangan Antar Muka..... | IV-18 |
| 4.3.1.2 Perancangan Data..... | IV-19 |
| 4.3.2 Kebutuhan Sistem..... | IV-20 |
| 4.3.3 Diagram <i>Sequence</i> | IV-20 |
| 4.4 Fase Kontruksi..... | IV-22 |
| 4.4.1 Diagram Kelas | IV-23 |
| 4.4.2 Implementasi | IV-23 |
| 4.4.2.1 Implementasi Kelas..... | IV-24 |
| 4.4.2.2 Implementasi Antarmuka..... | IV-26 |
| 4.5 Fase Transisi..... | IV-27 |
| 4.5.1 Pemodelan Bisnis | IV-27 |
| 4.5.2 Lingkungan Pengujian..... | IV-27 |
| 4.5.3 Rencana Pengujian | IV-28 |
| 4.5.4 Kasus Uji | IV-29 |
| BAB V HASIL ANALISA DAN PENELITIAN | V-1 |
| 5.1 Pendahuluan | V-1 |
| 5.2 Data Hasil Pengujian | V-1 |
| 5.2.1 Hasil Pengujian pada Data Uji Kesalahan Pengetikan Kata Dasar Menggunakan Algoritma <i>Soundex Similarity</i> | V-1 |
| 5.2.2 Hasil Penhgujian pada Data Uji Kesalahan Pengetikan Kata Dasar Menggunakan Algoritma Damerau-Levenshtein Distance..... | V-3 |
| 5.3 Analisis Hasil Pengujian..... | V-4 |
| 5.4 Kesimpulan..... | V-6 |
| BAB VI KESIMPULAN DAN SARAN | VI-1 |
| 6.1 Pendahuluan | VI-1 |
| 6.2 Kesimpulan..... | VI-1 |
| 6.3 Saran | VI-2 |
| DAFTAR PUSTAKA | xvi |
| LAMPIRAN | xviii |

DAFTAR TABEL

| | | |
|--------------|--|--------|
| Tabel II-1. | Kata Yang Salah & Kemungkinan Data | II-2 |
| Tabel II-2. | Matriks Pencarian Nilai Jarak <i>Levensthein Distance</i> | II-5 |
| Tabel II-3. | Matriks Pencarian Nilai Jarak <i>Damerau-Levensthein Distance</i> | II-9 |
| Tabel II-4. | Tabel Kode <i>Soundex Similarity</i> | II-10 |
| Tabel III-1. | Rancangan Tabel Hasil Koreksi Program Penelitian Menggunakan Algoritma <i>Damerau-Levenshtein Distance</i> dan <i>Soundex Similarity</i> | III-5 |
| Tabel III-2. | Hasil Ketepatan Koreksi Percobaan Konfigurasi I pada Algoritma <i>Damerau-Levenshtein Distance</i> dan <i>Soundex Similarity</i> | III-7 |
| Tabel III-3. | Hasil Perbandingan Akurasi Ketepatan Koreksi Percobaan | III-7 |
| Tabel III-4. | Tabel <i>Relational Unified Process (RUP)</i> | III-8 |
| Tabel III-5. | Tabel <i>Work Breakdown Structure (WBS)</i> Dari Penelitian Yang Akan Dilakukan | III-16 |
| Tabel IV-1. | Kebutuhan Fungsional..... | IV-4 |
| Tabel IV-2 | Kebutuhan Non-Fungsional..... | IV-4 |
| Tabel IV-3 | Contoh Proses Preprocessing..... | IV-7 |
| Tabel IV-4 | Hasil Pencocokan Kata..... | IV-7 |
| Tabel IV-5 | Pengkodean soundex dengan panjang karakter lebih dari 4 | IV-8 |
| Tabel IV-6. | Pengkodean Soundex karakter huruf yang sama dan Berhimpitan..... | IV-9 |
| Tabel IV-7 | Pengkodean Kata Typo dengan Kata yang memiliki kode yang sama..... | IV-9 |
| Tabel IV-8 | Matriks Langkah 2 dan 3..... | IV-10 |
| Tabel IV-9 | Matriks Langkah 2 dan 3..... | IV-11 |
| Tabel IV-10 | Matriks Langkah 2 dan 3..... | IV-12 |
| Tabel IV-11 | Jarak Beberapa istilah dari Tahap 4 – 7..... | IV-12 |
| Tabel IV-12 | Definisi Aktor..... | IV-14 |
| Tabel IV-13 | Definisi Use Case..... | IV-15 |
| Tabel IV-14 | Skenario <i>Use Case</i> Koreksi Ejaan..... | IV-15 |
| Tabel IV-15 | Implementasi Kelas..... | IV-24 |
| Tabel IV-16 | Rencana Pengujian <i>Use Case</i> | IV-28 |
| Tabel IV-17 | Pengujian <i>Use Case</i> | IV-30 |
| Tabel V-1. | Data Hasil Percobaan Konfigurasi I pada Algoritma <i>Soundex Similarity</i> .. | V-1 |

| | | |
|-----------|---|-----|
| Tabel V-2 | Data Hasil Percobaan Konfigurasi I pada Algoritma <i>Damerau-Levenshtein Distance</i> | V-3 |
| Tabel V-3 | Hasil Ketepatan Koreksi Percobaan Konfigurasi I pada Algoritma <i>Damerau-Levenshtein Distance</i> dan <i>Soundex Similarity</i> | V-5 |
| Tabel V-4 | Hasil Perbandingan Akurasi Ketepatan Koreksi Percobaan Konfigurasi I pada Algoritma <i>Damerau-Levenshtein Distance</i> dan <i>Soundex Similarity</i> | V-5 |

DAFTAR GAMBAR

| | |
|--|-------|
| Gambar II-1. Arsitektur <i>Relational Unified Process</i> (RUP) Penelitian Yang Akan Dilakukan | II-12 |
| Gambar III-1. Tahapan Pengujian Penelitian | III-6 |
| Gambar IV-1. Contoh Data Kamus Soundex Beserta Kode..... | IV-5 |
| Gambar IV-2. Diagram <i>Use Case</i> Perangkat Lunak Penelitian..... | IV-14 |
| Gambar IV-3. Diagram Aktivitas..... | IV-17 |
| Gambar IV-4. Rancangan Antarmuka..... | IV-19 |
| Gambar IV-5. <i>Sequence Diagram Load File</i> | IV-21 |
| Gambar IV-6. <i>Sequence Diagram</i> Pengoreksian Kata..... | IV-22 |
| Gambar IV-7. Diagram Kelas..... | IV-23 |
| Gambar IV-8. Tampilan Antarmuka Perangkat Lunak Sebelum Proses Pengoreksian.... | IV-26 |
| Gambar IV-9. Tampilan Antarmuka Perangkat Lunak Setelah Proses Pengoreksian..... | IV-27 |

DAFTAR LAMPIRAN

1. Source Code Program L-2

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan ini diuraikan mengenai latar belakang masalah, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, sistematika penulisan, serta kesimpulan.

1.2 Latar Belakang Masalah

Bahasa merupakan salah satu komponen terbesar dalam kehidupan manusia yang dapat diekspresikan melalui lisan atau tulisan. Sebagai teks, bahasa menjadi elemen penting pada penulisan dokumen. Kesalahan ejaan atau penulisan pada *editor* teks merupakan hal yang cukup sering terjadi, khususnya pada *word editor*. Beberapa kata memiliki kemiripan dalam susunan karakter sehingga dapat terjadi kesalahan dalam penulisannya. Kesalahan tersebut biasanya disebabkan oleh dua hal diantaranya kata tersebut tidak terdapat di dalam kamus *database* atau kelalaian pengguna. Untuk membantu proses pengoreksian kesalahan penulisan kata dibutuhkan sebuah perangkat lunak yang mampu membantu pekerjaan untuk menyunting atau memperbaiki teks berita, naskah tulisan, maupun skripsi. Apabila dalam pengoreksian ejaan kata ditemukan ejaan kata yang kurang tepat maka harus dilakukan pencarian kemungkinan kata yang sesuai. Pada proses pencarian kemungkinan kata tersebut, dibutuhkan suatu pendekatan pencarian *string* khusus yaitu menggunakan algoritma *approximate string matching*, salah satunya *damerau Levenshtein Distance* serta *Phonetic string matching*, salah satunya algoritma *soundex similarity* (Madaharsa & Adiwidya, 2009).

Berdasarkan penelitian yang dilakukan oleh Tusty et al (2017), Algoritma *damerau Levenshtein Distance* dapat digunakan untuk mendeteksi kesalahan ejaan dan penentuan rekomendasi koreksi kata dengan tingkat ketepatan yang cukup tinggi. Penelitian ini dilakukan dengan cara mengecek kesalahan pengetikan terhadap dokumen jurnal berbahasa Indonesia (JTIK) menggunakan metode lookup dictionary lalu dilanjutkan dengan memilih rekomendasi kata menggunakan algoritma *damerau Levenshtein Distance*. Hasil penelitian yang dilakukan oleh Tusty menunjukkan bahwa algoritma *damerau-Levenshtein Distance* dapat menghasilkan hasil akurasi yang baik dengan hasil nilai presisi sebesar 0.78 dan recall sebesar 1.

Pada penelitian lainnya yang dilakukan oleh Ahmad Pahdi(2016), Algoritma *Soundex Similarity* juga dapat digunakan untuk mendeteksi kesalahan ejaan dan penentuan rekomendasi koreksi kata pada istilah computer dengan tingkat ketepatan yang baik. Penelitian ini dilakukan dengan cara menguji masing masing algoritma yaitu *damerau Levenshtein Distance* dan *Soundex Similarity* untuk mengecek kesalahan dan rekomendasi dalam istilah komputer. Hasil penelitian yang dilakukan oleh Pahdi menunjukkan bahwa algoritma *soundex similarity* dapat menghasilkan hasil akurasi yang baik dengan hasil nilai akurasi sebesar 74% dibanding dengan algoritma *damerau-Levenshtein Distance* yang hanya menghasilkan akurasi sebesar 70%.

Walaupun penelitian terkait telah dilakukan, Perbandingan kedua metode dalam mengoreksi dan merekomendasikan ejaan kata pada secara otomatis belum pernah dilakukan sebelumnya sehingga belum diketahui secara jelas metode mana

yang lebih baik. Oleh karena itu, Penelitian ini akan melakukan perbandingan antara algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity* pada pengoreksian ejaan kata otomatis agar dapat mengetahui metode terbaik yang dapat dipakai dalam mengoreksi ejaan kata secara otomatis.

1.3 Rumusan Masalah

Rumusan masalah dari penelitian yang dilakukan adalah bagaimana hasil pengoreksian ejaan kata otomatis dari algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity* ?. Untuk menjawab rumusan masalah, maka diuraikan pertanyaan sebagai berikut :

1. Bagaimana membangun perangkat lunak berdasarkan algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity* sebagai alat untuk mengoreksi ejaan kata yang salah ?
2. Bagaimana hasil performansi akurasi algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity* dalam mengoreksi ejaan kata yang salah?

1.4 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah sebagai berikut :

1. Mengetahui analisis dan rancangan perangkat lunak berdasarkan algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity* sebagai alat penelitian dalam mengoreksi ejaan kata yang salah
2. Mengetahui metode mana yang lebih optimal dalam mengoreksi ejaan kata antara algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity*.

1.5 Manfaat Penelitian

Manfaat dari penelitian yang dilakukan adalah sebagai berikut :

1. Dapat memahami proses analisis dan rancangan perangkat lunak berbasis algoritma *Damerau-Levenshtein Distance* dan *Soundex Similarity* dalam mengoreksi ejaan kata yang salah.
2. Dapat mengetahui metode mana yang lebih optimal dalam dalam mengoreksi ejaan kata yang salah.
3. Mampu menerapkan metode pengoreksian ejaan kata otomatis dalam implementasi perangkat lunak.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini antara lain :

1. Metode penelitian hanya berfokus pada algoritma pengoreksian kata yaitu algoritma *damerau-Levenshtein Distance* dan *soundex similarity*
2. Data yang dimasukkan berupa teks berbahasa Indonesia.
3. Teks yang dimasukkan berupa karakter alfabet.
4. Teks yang diinputkan berupa kata dasar tanpa imbuhan.
5. Masukkan berupa berkas teks (.txt) dan input dari *keyboard*.
6. Batas maksimum kata yang bisa dimasukkan hanya sebanyak 100 karakter.

1.7 Sistematika Penulisan

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian seperti definisi pengoreksian ejaan kata (*spelling correction*), metode pencarian jarak terkecil (*Approximate String Matching*), Algoritma *Damerau-Levenshtein Distance*, algoritma *Soundex Similarity* dan metode pengembangan perangkat lunak yang akan diterapkan di dalam penelitian.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

1.8 Kesimpulan

Penelitian yang akan dilakukan berfokus untuk mengetahui performa dari algoritma *Damerau-Levenshtein Distance* dan algoritma *Soundex Similarity* dalam

mengoreksi kesalahan penulisan kata berbahasa Indonesia secara otomatis. Perbandingan antara algoritma *Damerau-Levenshtein Distance* dan algoritma *Soundex Similarity* dilakukan untuk mengetahui algoritma mana yang lebih efektif dalam mengoreksi ejaan kata otomatis. Dengan mengetahui algoritma mana yang lebih efektif diharapkan akan mampu memberikan kontribusi positif terhadap penelitian sejenis di masa yang akan datang.

DAFTAR PUSTAKA

- Al-khamaiseh, K., & Alshagarin, S. (2014). OPEN ACCESS A Survey of String Matching Algorithms, *4*(7), 144–156.
- Ali, A. (2011). Textual Similarity.
- Angeles, L. (2014). of Evaluation, (September 2012).
<https://doi.org/10.1177/1098214011426594>
- Braddley, M. O., & Fachrurrozi, M. (n.d.). Pengoreksian Ejaan Kata Berbahasa Indonesia Menggunakan Algoritma Levenshtein Distance.
- Fachrurrozi, M., & Manik, A. A. (n.d.). Perbaikan Ejaan Kata pada Dokumen Bahasa Indonesia dengan Metode Cosine Similarity.
- Junaidi, J., & Jambi, U. (2015). Prosedur Uji Chi-Square, (June).
<https://doi.org/10.13140/RG.2.1.2797.8400>
- Kruchten, P. (2003). *The Rational Unified Process An Introduction*.
<https://doi.org/10.1109/ICSE.2002.146346>
- Madaharsa, B., & Adiwidya, D. (2009). Algoritma Levenshtein Dalam Pendekatan Approximate.
- Martins, B., & Silva, M. J. (2014). Spelling Correction for Search Engine Queries, (May). <https://doi.org/10.1007/978-3-540-30228-5>

Sutisna, U. (2010). Koreksi Ejaan Query Bahasa Indonesia Menggunakan Algoritme Damerau Levenshtein, *15*(2), 25–29. Retrieved from <http://repository.ipb.ac.id/handle/123456789/30608>