

Analisa Perbandingan Algoritma *Damerau-Levensthein Distance* dan *Soundex Similarity* Pada Pengkoreksian Ejaan Kata Otomatis

Abson Hadi, Muhammad Fachrurrozi, Purwita
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
Email: obetsobets@gmail.com

Abstrak— Pada penelitian ini, Algoritma *Damerau-Levensthein Distance* dan *Soundex Similarity* digunakan untuk membangun perangkat lunak untuk mengoreksi ejaan kata yang salah. Pada tahapan pengujian performansi akurasi algoritma *Damerau-Levensthein Distance* dan *Soundex Similarity* dalam mengoreksi ejaan kata diuji dan memiliki performa optimal.

Kata kunci— *Damerau-Levensthein Distance*, *Soundex Similarity*, ejaan.

I. PENDAHULUAN

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 paling sering terjadi, khususnya di dalam *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 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 naskah tulisan, skripsi maupun teks berita. 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, diperlukan suatu pendekatan pencarian string khusus yaitu dengan menggunakan algoritma *approximate string matching*, salah satunya *damerau levensthein distance* serta *Phonetic string matching*, salah satunya *algoritma soundex similarity*[1].

Algoritma *damerau levensthein distance* dapat digunakan untuk mendeteksi kesalahan ejaan dan penentuan rekomendasi koreksi kata dengan tingkat ketepatan yang cukup tinggi[2]. 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 levensthein distance*. Hasil penelitian yang dilakukan oleh tusty menunjukkan bahwa algoritma *damerau-levensthein distance* dapat menghasilkan hasil akurasi yang baik dengan hasil nilai presisi sebesar 0.78 dan recall sebesar 1.

Algoritma *Soundex Similarity* juga dapat digunakan untuk mendeteksi kesalahan ejaan dan penentuan rekomendasi koreksi kata pada istilah komputer dengan tingkat ketepatan

yang baik[3]. Penelitian ini dilakukan dengan cara menguji masing masing algoritma yaitu *damerau levensthein 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-levensthein 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-Levensthein Distance* dan *Soundex Similarity* pada pengoreksian ejaan kata otomatis agar dapat mengetahui metode terbaik yang dapat dipakai dalam mengoreksi ejaan kata secara otomatis. Penelitian ini bertujuan untuk Mengetahui analisis dan rancangan perangkat lunak berdasarkan algoritma *Damerau-Levensthein Distance* dan *Soundex Similarity* sebagai alat penelitian dalam mengoreksi ejaan kata yang salah dan mengetahui metode mana yang lebih optimal dalam mengoreksi ejaan kata antara algoritma *Damerau-Levensthein Distance* dan *Soundex Similarity*.

II. KAJIAN LITERATUR

A. Kesalahan Ejaan (*Spelling Error*)

Kesalahan ejaan (*spelling error*) merupakan keadaan di mana terjadi kesalahan penulisan ejaan atau susunan kata. Berdasarkan sejarahnya, hal ini disebabkan kesalahan penulisan secara manual, namun saat ini sering terjadi pada proses pengetikan dengan bantuan komputer baik karena kelalaian pengguna atau ketidaktahuan pengguna tentang pengejaan tulisan yang benar. Terdapat dua metode yang dapat digunakan untuk mendeteksi kesalahan ejaan yaitu *N-gram analysis* dan *dictionary lookup*. Metode yang digunakan pada penelitian ini adalah *dictionary lookup* yaitu mencari kata-kata yang salah dengan cara membandingkan langsung dengan kamus.

B. *Spelling Correction*

Spelling Correction adalah sistem yang mampu mengoreksi kesalahan ejaan kata. Kesalahan dapat terjadi karena kurang atau lebihnya karakter atau karena terdapat karakter yang tidak disesuaikan[1].

Spelling correction merupakan proses yang dilakukan setelah proses deteksi kesalahan ejaan selesai. Setelah menentukan suatu kata terdapat kesalahan ejaan, pada proses ini dilakukan pencarian kata sebagai kandidat untuk mengoreksi kata yang salah ejaan tersebut.

Tabel 1 Kata Yang Salah Dan Kemungkinan Kata

Kata yang salah	Kemungkinan Kata	Penyebab Kesalahan
Kacua	Kacau	Kesalahan urutan huruf 'a' dan 'u'
	Kaca	Terdapat huruf 'u'
Pana	Panas	Kurang huruf 's'
Temn	Teman	Kurang huruf 'a'
	Tema	Kesalahan huruf 'n' dan yang seharusnya adalah huruf 'a'
Auk	Aku	Kesalahan urutan huruf 'u' dan 'k'

C. Approximate String Matching

Approximate Matching hingga saat ini sering diaplikasikan sebagai metode pencocokan string. Hal ini disebabkan karena kualitas teks yang rendah, heterogenitas basis data, kesalahan ejaan kata, mencari nama asing dan ketidakpastian[4].

Approximate String Matching adalah teknik untuk pencocokkan pola pada string dengan cara pendekatan. Terdapat tiga operasi dasar pada pendekatan tersebut yaitu operasi penghapusan, operasi penambahan, dan penggantian[1]. Metode ini juga dapat diterapkan ke dalam beberapa algoritma dua diantaranya yaitu *Damerau-Levenshtein Distance* dan *Soundex Similarity*.

D. Praproses

Teks praproses adalah bagian penting dari setiap sistem *Natural Language Processing*, karena karakter, kata, dan kalimat yang diidentifikasi pada tahap ini adalah unit dasar yang akan diteruskan ke semua tahap pemrosesan lebih lanjut, seperti analisis morfologi, *part of speech taggers* dan pengolahan data dengan *machine learning*[5]. Praproses adalah proses awal mengelola data sebelum pengolahan data dilakukan. Tahapan-tahapan yang dilakukan pada praproses diantaranya yaitu *casefolding*, *tokenizing*, dan *stemming*.

1) *Casefolding* adalah proses penyamaan *case* dalam sebuah dokumen, dari huruf besar ke huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dianggap *delimiter*. Sebagai contoh *casefolding*, kalimat 'Ria Membaca Buku di Perpustakaan' menjadi 'ria membaca buku'.

2) *Tokenizing* adalah proses memecah kalimat menjadi kumpulan kata, contohnya kalimat 'ria membaca buku', menjadi kumpulan kata 'ria', 'membaca', 'buku'.

3) *Stopword Removal* adalah proses menghapus kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna, contoh II-3 *stopwords* dalam bahasa Inggris diantaranya 'of', 'the', sedangkan dalam bahasa Indonesia diantaranya 'yang', 'ke'.

4) *Stemming* adalah proses untuk menemukan kata dasar dari sebuah kata dengan menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan. *Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi Bahasa Indonesia yang baik dan benar. Sebagai contoh *stemming*, kata 'membaca' menjadi 'baca' dengan menghilangkan imbuhan awal kata (*prefixes*) 'mem'.

E. Metode Pengumpulan data

Data dibuat secara manual dengan memasukkan data berupa teks yang kemudian data tersebut disimpan dengan format .txt

III. METODOLOGI PENELITIAN

Bagian ini memaparkan lebih rinci mengenai tahapan penelitian yang akan dilakukan. Kerangka penelitian yang akan dipaparkan meliputi kerangka kerja, kriteria pengujian, format data pengujian, alat yang digunakan dalam pelaksanaan penelitian, pengujian penelitian, analisis hasil pengujian dan sintesis kesimpulan.

A. Kerangka Kerja

1. Preprocessing

Preprocessing merupakan suatu tahap yang dilakukan sebelum proses pengolahan data dikerjakan. Proses utamanya adalah penerapan algoritma *Damerau-Levenshtein Distance* dan algoritma *Soundex Similarity*. Pada tahap ini dokumen akan melalui empat *preprocessing* yaitu *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*.

2. Algoritma Damerau-Levenshtein Distance

Algoritma *Damerau-Levenshtein Distance* merupakan algoritma pengembangan dari algoritma *Damerau-Levenshtein*. *Damerau-Levenshtein Distance* menentukan jumlah operasi minimum yang diperlukan untuk mengubah suatu string atau kata yang salah ke string benar melalui empat jenis operasi penggantian, penghapusan, penambahan, dan penukaran posisi karakter. Operasi algoritma *Damerau-Levenshtein Distance* dimulai dengan menginisialisasi nilai n sebagai panjang karakter dari kata kunci, dan m sebagai panjang karakter yang akan diukur kedekatannya. Setelah di dapat nilai n dan m , matriks $n+1 \times m+1$ akan dibuat untuk mengukur jarak edit kedekatan antara kata kunci dan kata target dengan menggunakan rumus algoritma *Damerau-Levenshtein Distance*. Selanjutnya, jarak maksimum edit ditentukan untuk memilih rekomendasi kata yang akan direkomendasikan berdasarkan jarak edit yang berada dibawah ambang batas jarak maksimum yang telah ditentukan sebelumnya.

3. Algoritma Soundex Similarity

Tahapan *spelling correction* menggunakan algoritma *soundex similarity* dilakukan berdasarkan kemiripan bunyi pada kata kunci dan kata target. Sebelumnya, kata target telah ditetapkan dalam kode yang telah dibuat yang selanjutnya disebut kode *soundex*. Tahapan untuk mengecek ejaan menggunakan algoritma *soundex* dibagi kedalam 3 tahapan. Tahapan pertama dimulai dengan melakukan *uppercase* pada seluruh kata dalam teks dan selanjutnya melakukan pembuangan huruf vocal, h, w, y dan huruf yang sama dalam satu kata. Tahapan kedua, Penggantian huruf dengan kode numerik *soundex* yang telah ditetapkan kecuali huruf pertama pada kata yang akan diproses. Tahapan ketiga, Pengambilan 4 digit kode pertama yang telah terbentuk dari tahapan kedua sehingga menjadi kode *soundex*. Setelah kode *soundex* didapatkan, pencocokan kode yang telah terbentuk pada tahap tiga dan kode *soundex* akan dilakukan. Jika terdapat kecocokan antara kode *soundex* dan kode yang telah dihasilkan pada tahap tiga, maka kata yang sesuai kode *soundex* yang dimaksud akan disarankan sebagai rekomendasi koreksi kata yang benar.

4. Analisa Hasil Pengujian

Pengukuran performansi dari metode *damerau-levensthein distance* dan *soundex similarity* akan dihitung berdasarkan nilai persentase akurasi. Nilai persentase akurasi dihitung berdasarkan jumlah ketepatan koreksi kata dibagi total kata yang dikoreksi yang telah didapat dari *spelling correction* menggunakan kedua metode dalam penelitian.

B. Kriteria Pengujian

Pada tahapan pengujian, beberapa kriteria perlu ditetapkan agar dapat dilakukan pengujian untuk mendapatkan hasil yang lebih optimal sebelum pengujian dilakukan. Penetapan kriteria pada penelitian yang akan dilakukan perlu dilakukan pada metode yaitu *damerau-levensthein distance*. Pada metode *damerau-levensthein distance* penetapan kriteria dilakukan pada tahapan penentuan jarak edit maksimal. Adapun kriteria yang perlu ditetapkan merupakan nilai toleransi kedekatan antara kata kunci dan kata target. Pada penelitian ini, Nilai toleransi kedekatan antara kata kunci dan kata target yang ditetapkan sebesar 40%.

C. Format Data Pengujian

Hasil pengujian *spelling correction* menggunakan kedua metode dengan *confusion matrix* pada tiap pengujian akan digambarkan dalam Tabel II.

Tabel 2 Rancangan Tabel Hasil Koreksi Program Penelitian Menggunakan Algoritma Damerau Levensthein Distance dan Soundex Similarity

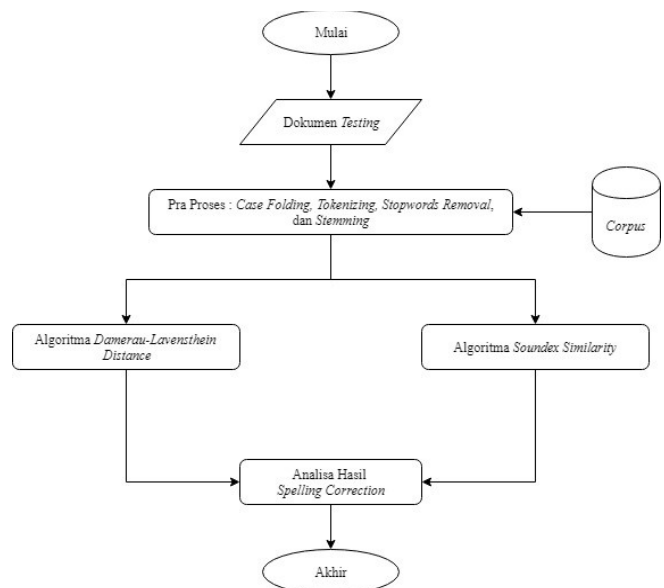
No	Kata Target	Kata Kunci	Kata Koreksi	Validasi
1				
..				
N				

D. Alat yang Digunakan dalam pengujian

Dalam menentukan nilai performansi dari hasil pengujian tiap metode yang ditunjukkan, dalam hal ini algoritma *Damerau-Levensthein Distance* dan *Soundex Similarity*, maka dibutuhkan sebuah alat penelitian yang dapat menghitung

nilai performansi yang dibutuhkan. Oleh karena itu, peneliti akan mengembangkan sebuah perangkat lunak yang dapat menghitung nilai akurasi berdasarkan algoritma *Damerau-Levensthein Distance* dan *Soundex Similarity* agar dapat mencapai kesimpulan akhir dari penelitian.

E. Penelitian



Gambar 1. Tahapan Pengujian Penelitian

F. Analisis Hasil Pengujian

Berikut adalah hasil ketepatan pengujian perangkat lunak yang telah dilakukan menggunakan konfigurasi percobaan data uji kesalahan pengetikan pada kata dasar dengan algoritma *damerau-levensthein distance* dan *soundex similarity*:

Tabel 3 Hasil Ketepatan Koreksi Percobaan Konfigurasi I Pada Algoritma Damerau Levensthein Distance dan Soundex Similarity

	Algoritma Damerau-Levensthein Distance	Algoritma Soundex Similarity
Tepat		
Tidak		

Berikut adalah perbandingan hasil analisis pengujian perangkat lunak yang telah dilakukan menggunakan konfigurasi percobaan data uji kesalahan pengetikan pada kata dasar dengan algoritma *damerau-levensthein distance* dan *soundex similarity*:

Tabel 4 Hasil Perbandingan Akurasi Ketepatan Koreksi Percobaan Konfigurasi I Pada Algoritma Damerau Levensthein Distance dan Soundex Similarity

	Algoritma Damerau-Levensthein Distance	Algoritma Soundex Similarity
Akruasi		

IV. HASIL DAN ANALISIS PENELITIAN

A. Data Hasil Pengujian

Hasil Pengujian pada Data Uji Kesalahan Pengetikan Kata Dasar Menggunakan Algoritma Soundex Similarity Authors and Affiliations

Berikut adalah hasil pengujian perangkat lunak yang telah dibuat menggunakan konfigurasi percobaan data uji kesalahan pengetikan pada kata dasar menggunakan algoritma *soundex similarity*:

Tabel 5 Data Hasil Percobaan Konfigurasi I Pada Algoritma Soundex Similarity

No	Kata Kunci	Kata Target	Hasil Koreksi	Validasi
1	Ikn	Ikan	Ikan	Benar
2	Pus	Puasa	Paus	Salah
3	Adalh	Adalah	Adalah	Benar
4	Hwan	Hewan	Hanya	Salah
5	Maali	Mamalia	Malah	Salah
6	Yg	Yang	Yg	Salah
7	Hdp	Hidup	Hdp	Salah
8	Satelat	Satelit	Satelit	Benar
9	Almi	Alami	Alami	Benar
10	Dr	Dari	Dr	Salah
11	Plant	Planet	Pelintir	Salah
12	Bum	Bumi	Bumi	Benar
13	Tlur	Telur	Telur	Benar
14	Merupakn	Merupakan	Merupakn	Salah
15	Sumbr	Sumber	Sembarang	Salah
16	Potein	Protein	Patin	Salah
17	Heani	Hewani	Hanya	Salah
18	Mrah	Murah	Marah	Salah
19	Satelat	Satelit	Satelit	Benar
20	Almi	Alami	Alami	Benar
21	Ekonmi	Ekonomi	Ekonomis	Salah
22	Faktr	Faktor	Faktor	Benar
23	Fgur	Figur	Figur	Benar
24	Gand	Ganda	Ganti	Salah
25	His	Hias	Hijau	Salah
26	Janwari	Januari	Jemur	Salah
27	Jepng	Jepang	Jepang	Benar
28	Merka	Mereka	Mereka	Benar
29	Seikat	Serikat	Sekte	Salah
30	Adgan	Adegan	Adakan	Salah
31	Absn	Absen	Absensi	Salah
32	Adipr	Adipura	Adipura	Benar
33	Amatr	Amatir	Amatir	Benar
34	Berintas	Barantas	Berantas	Benar
35	Dendung	Dendang	Dandang	Salah
36	Gandim	Gandum	Gandum	Benar
37	Helikopter	Helikopter	Helikopter	Benar
38	Imajinase	Imajinasi	Imajinasi	Benar
39	Jangkr	Jangkar	Jangkung	Salah
40	Kelilang	Keliling	Keliling	Benar
41	Kelanu	Kelana	Kelam	Salah
42	Kelompk	Kelompok	Kelompok	Benar
43	Kinrja	Kinerja	Kinerja	Benar
44	Kmbinasi	Kombinasi	Kambium	Salah
45	Kurukulum	Kurukulum	Kurikulum	Benar
46	Lustrik	Listrik	Listrik	Benar

47	Malapetika	Malapetaka	Malapetaka	Benar
48	Navigasu	Navigasi	Navigasi	Benar
49	Olimpyade	Olimpiade	Olimpiade	Benar
50	Operasuonal	Operasiaonal	Operasional	Benar

Hasil Pengujian pada Data Uji Kesalahan Pengetikan Kata Dasar Menggunakan Algoritma Damerau Levensthein Distance

Berikut adalah hasil pengujian perangkat lunak yang telah dibuat menggunakan konfigurasi percobaan data uji kesalahan pengetikan pada kata dasar menggunakan algoritma *damerau-levensthein distance*:

Tabel 6 Data Hasil Percobaan Konfigurasi I Pada Algoritma Damerau Levensthein Distance

No	Kata Kunci	Kata Target	Hasil Koreksi	Validasi
1	Ikn	Ikan	Ikan	Benar
2	Pus	Puasa	Pos	Salah
3	Adalh	Adalah	Adalah	Benar
4	Hwan	Hewan	Hewan	Benar
5	Maali	Mamalia	Maali	Salah
6	Yg	Yang	Ya	Salah
7	Hdp	Hidup	Hdp	Salah
9	Bln	Bulan	Bln	Salah
10	Satelat	Satelit	Satelit	Benar
11	Almi	Alami	Alami	Benar
12	Dr	Dari	Di	Salah
13	Plant	Planet	Plant	Salah
14	Bum	Bumi	Bumi	Benar
15	Tlur	Telur	Telur	Benar
16	Merupakn	Merupakan	Merupakn	Salah
17	Sumbr	Sumber	Sumber	Benar
18	Potein	Protein	Protein	Benar
19	Heani	Hewani	Hewani	Benar
20	Mrah	Murah	Marah	Salah
21	Ekonmi	Ekonomi	Ekonomi	Benar
22	Faktr	Faktor	Fakir	Salah
23	Fgur	Figur	Figur	Benar
24	Gand	Ganda	Ganda	Benar
25	His	Hias	Hias	Benar
26	Janwari	Januari	Januari	Benar
27	Jepng	Jepang	Jepang	Benar
28	Merka	Mereka	Mereka	Benar
29	Seikat	Serikat	Serikat	Benar
30	Adgan	Adegan	Adegan	Benar
31	Absn	Absen	Aben	Salah
32	Adipr	Adipura	Adipr	Salah
33	Amatr	Amatir	Amat	Salah
34	Berintas	Barantas	Berantas	Benar
35	Dendung	Dendang	Dendang	Benar
36	Gandim	Gandum	Gandin	Salah
37	Helikopter	Helikopter	Helikopter	Benar
38	Imajinase	Imajinasi	Imajinasi	Benar
39	Jangkr	Jangkar	Jangka	Salah
40	Kelilang	Keliling	Keliling	Benar
41	Kelanu	Kelana	Kelana	Benar
42	Kelompk	Kelompok	Kelompok	Benar
43	Kinrja	Kinerja	Kinerja	Benar
44	Kmbinasi	Kombinasi	Kombinasi	Benar
45	Kurukulum	Kurukulum	Kurukulum	Benar

46	Lustrik	Listrik	Listrik	Benar
47	Malapetika	Malapetaka	Malapetaka	Benar
48	Navigasu	Navigasi	Navigasi	Benar
49	Olimpyade	Olimpiade	Olimpiade	Benar
50	Operasuonal	Operasiaonal	Operasiaonal	Benar

B. Analisis Hasil Pengujian

Berikut adalah hasil ketepatan pengujian perangkat lunak yang telah dilakukan menggunakan konfigurasi percobaan data uji kesalahan pengetikan pada kata dasar dengan algoritma *damerau-levensthein distance* dan *soundex similarity*:

Tabel 7 Hasil Ketepatan Koreksi Percobaan Konfigurasi I Pada Algoritma Damerau Levensthein Distance dan Soundex Similarity

	Algoritma Damerau-Levensthein Distance	Algoritma Soundex Similarity
Tepat	36	25
Tidak	16	25
Total	50	50

Berikut adalah perbandingan hasil analisis pengujian perangkat lunak yang telah dilakukan menggunakan konfigurasi percobaan data uji kesalahan pengetikan pada kata dasar dengan algoritma *damerau-levensthein distance* dan *soundex similarity*:

Tabel 8 Hasil Perbandingan Akurasi Ketepatan Koreksi Percobaan Konfigurasi I Pada Algoritma Damerau Levensthein Distance dan Soundex Similarity

	Algoritma Damerau-Levensthein Distance	Algoritma Soundex Similarity
Akruasi	68%	50%

Tabel 8, menunjukkan bahwa algoritma *damerau levensthein distance* dapat lebih baik mengoreksi ejaan kata secara otomatis dibandingkan dengan algoritma *soundex similarity*. Hasil akurasi tinggi dari algoritma *damerau levensthein distance* didapat dikarenakan algoritma *damerau levensthein distance* cenderung fleksibel dan semakin baik dalam mengoreksi ejaan kata secara otomatis jika *corvus* yang dipakai semakin lengkap, dengan kata lain semakin banyak jumlah kata yang tersedia dalam *corvus* maka semakin baik pula algoritma *damerau levensthein* dapat mencari rekomendasi kata dikarenakan semakin banyak kemungkinan kata yang masuk ke dalam jarak maksimum.

Sedangkan hasil akurasi rendah yang didapat oleh algoritma *soundex* disebabkan oleh beberapa hal. Pertama, tidak terdapatnya kode *fonetis* khusus yang dapat digunakan dalam mengkodekan kata Bahasa Indonesia. Penggunaan kode *fonetis* yang tersedia sekarang lebih dikhususkan untuk pengkodean kata ejaan dalam Bahasa Inggris, sehingga diperlukan aturan kode *fonetis* yang lebih spesifik untuk meningkatkan akurasi dari algoritma *soundex similarity* dalam mengoreksi ejaan kata berbahasa Indonesia. Kedua, Terbatasnya *corvus* kode *soundex* untuk kata dalam Bahasa Indonesia yang menyebabkan algoritma *soundex similarity* terkadang tidak dapat merekomendasikan kata target yang dimaksud dengan tepat. Ketiga, Permasalahan pengkodean kata yang memiliki 2 suku kata yang terlalu pendek terkadang menyebabkan algoritma *soundex similarity* tidak dapat mengoreksi dan merekomendasikan kata dengan tepat.

V. KESIMPULAN

Berdasarkan hasil percobaan yang telah dilakukan, dapat disimpulkan bahwa:

1. Algoritma *damerau-levensthein distance* lebih baik daripada *soundex similarity* dalam pengoreksian ejaan kata secara otomatis.
2. Berdasarkan pengujian yang telah dilakukan, algoritma *damerau-levensthein distance* mendapat hasil akurasi sebesar 68%, sedangkan algoritma *soundex similarity* mendapat akurasi sebesar 50%.

DAFTAR PUSTAKA

- [1] B. M. Adiwidya, "Algoritma Levenshtein Dalam Pendekatan Approximate String Matching," *Makal. IF3051 Strateg. Algoritm.*, 2009.
- [2] T. N. Maghfira, I. Cholissodin, and A. W. Widodo, "Deteksi kesalahan ejaan dan penentuan rekomendasi koreksi kata yang tepat pada dokumen jurnal JTIK menggunakan dictionary lookup dan damerau-levenshtein distance," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, p. 964X, 2017.
- [3] A. P.-S. Banjarbaru, "Koreksi Ejaan Istilah Komputer Berbasis Kombinasi Algoritma Damerau-Levenshtein dan Algoritma Soundex," *Speed-Sentra Penelit. Eng. dan Edukasi*, vol. 8, no. 2, 2017.
- [4] K. Al-Khamaiseh and S. ALShagarin, "A survey of string matching algorithms," *Int. J. Eng. Res. Appl.*, vol. 4, no. 7, pp. 144–156, 2014.
- [5] S. M. Weiss, N. Indurkha, T. Zhang, and F. Damerau, *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media, 2010.