

Perbandingan Metode Peter Norvig, Levenshtein *distance*, dan Damerau-Levenshtein *distance* : Tinjauan Literatur

Anggasta Tirta Adi Kusuma
Program Studi Informatika – Program Sarjana
Fakultas Teknologi Industri, Universitas Islam Indonesia
Yogyakarta, Indonesia
19523126@students.uii.ac.id

Chanifah Indah Ratnasari
Program Studi Informatika
Fakultas Teknologi Industri, Universitas Islam Indonesia
Yogyakarta, Indonesia
chanifah.indah@uui.ac.id

Abstract — *Teks merupakan media tulis yang digunakan untuk media komunikasi dan sumber dari segala informasi bagi kehidupan manusia. Media tulis tersebut bisa menjadi salah arti apabila terdapat Typographical error. Typographical error dapat diatasi dengan menggunakan alat bantu berupa sistem kesalahan ejaan atau spell checking. Makalah ini membahas mengenai Spell Checking Bahasa Indonesia dengan melakukan tinjauan literatur perbandingan Metode Peter Norvig, Levenshtein distance, dan Damerau-Levenshtein distance. Kajian literatur ini didapatkan melalui referensi yang tersebar dalam situs web seperti Google Scholar, Science Direct, dan Mendeley. Dengan melakukan pemilahan kata kunci yaitu, Spell checking, Bahasa Indonesia, Spelling Error, Peter Norvig, Levenshtein Distance, Damerau-Levenshtein Distance. Kajian literatur yang sudah dipilah dan dipilih ada 10 kajian literatur dengan jangka waktu dalam tahun dari 2017 sampai 2022. Kajian literatur tersebut dirangkum ke dalam tabel yang memiliki poin seperti peneliti, bahasan, tahun, metode, dan hasil. Studi literatur ini memiliki kesimpulan mengenai spell checking dimana pendeteksi kesalahan tidak selamanya bisa mendeteksi semua kesalahan yang ada dalam dokumen atau kalimat tertentu karena ada batasan-batasan di dalam korpus yang sudah disediakan tetapi dengan menggunakan perbandingan beberapa metode seperti Metode Peter Norvig, Levenshtein distance, dan Damerau-Levenshtein distance dapat mengurangi kesalahan penulisan.*

Keywords — *Spell checking, Bahasa Indonesia, Spelling Error, Peter Norvig, Levenshtein Distance, Damerau-Levenshtein Distance.*

I. PENDAHULUAN

Bahasa adalah salah satu komponen terpenting dalam kehidupan manusia saat ini, sebagai makhluk yang berakal dan bisa mengekspresikan segala hal yang dirasakan pada dirinya baik dengan mengucapkan secara lisan atau menulis tulisan. Ada berbagai macam bahasa yang dipakai masyarakat yang ada di seluruh belahan negara, salah satunya adalah Bahasa Indonesia. Bahasa Indonesia saat ini sudah dipakai lebih dari 222 juta orang yang dipakai dalam kehidupan sehari-hari[1]. Bahasa juga menjadi hal yang penting untuk penulisan baik secara formal atau tidak formal,

penulisan formal seperti jurnal ilmiah, dokumen akademik, dokumen hukum dan lain-lainnya[2]

Menulis tulisan atau sebuah dokumen yang penting harus memiliki keterampilan yang baik dikarenakan menulis adalah suatu aktivitas untuk menyampaikan pendapat yang tidak bisa secara langsung disampaikan bahkan tidak dapat diterima secara langsung oleh pihak yang dituju. Kesalahan dalam penulisan terkadang dapat membuat arti dari kata yang akan disampaikan, kesalahan dalam penulisan dapat diatasi dengan cara lebih teliti dalam melakukan penulisan atau dengan menggunakan alat bantu seperti *spell checking*.

Spell checking memiliki fungsi sebagai alat bantu untuk mengecek suatu ejaan yang salah berdasarkan bahasa dan metode tertentu yang digunakan pada proses *spell checking*, *spell checking* memiliki cara kerja yaitu pemindaian kata-kata pada kalimat, membandingkan kata yang salah dan kata yang benar menurut sumber data kumpulan kata yang sudah disiapkan untuk melakukan pengecekan pada kalimat yang salah. Diberikan sebuah saran kata yang sesuai sehingga penulis dapat memilih atau melakukan pengecekan pada kata mana yang terdapat sebuah kesalahan pada kalimat yang diujikan. Dalam beberapa kasus *spell checking* tidak 100% benar ini disebabkan ketidakakuratan kata yang terdapat pada sistem tersebut.

Pada penelitian ini akan dilakukan uji coba 3 metode yaitu, Peter Norvig berfungsi untuk mengubah satu jarak kata pada kata yang memiliki sebuah kesalahan akan digantikan menjadi dua kata. Levenshtein *distance*, dan Damerau-Levenshtein *distance*. Levenshtein *distance* sendiri menggunakan 3 operasi untuk menciptakan sebuah *string*. *Insertion* (penambahan karakter), *deletion* (penghapusan karakter), dan *substitution* (penggantian karakter). Damerau Levenshtein *distance* sendiri merupakan metode yang memiliki peningkatan dari metode sebelumnya yaitu Levenshtein *distance*, memiliki operasi yang sama hanya saja ada penambahan yaitu proses *transposisi* atau pertukaran posisi dua karakter.

Tujuan studi ini yaitu untuk mengetahui tingkat akurasi dari ke-tiga metode yang sudah disebutkan, untuk uji coba pada sistem *spell checking* untuk melakukan pendeteksian, pemberi saran kata yang tepat, serta melakukan pembenaran pada kalimat uji. Diharapkan dengan penelitian ini dapat dikembangkan dengan berbagai macam metode lainnya untuk perkembangan *spell checking* sehingga di dapat akurasi pembenaran kata mencapai tingkat akurasi yang terbaik.

II. METODE

Penelitian ini dilakukan untuk melakukan peninjauan literatur untuk perkembangan *spell checking* Bahasa Indonesia dengan melakukan perbandingan dari beberapa metode. Kata kunci di dalam penelitian ini adalah: (a) “*Spell Checking*”, (b) “Bahasa Indonesia”, (c) “*Spelling Error*”, (d) “Peter Norvig”, (e) “*Levenshtein Distance*”, (f) “*Damerau-Levenshtein Distance*”. Literatur yang dipilih di dapatkan dengan menggunakan **Google Scholar**, **Science Direct**, dan **Mendeley** dengan menggunakan kata kunci tersebut.

III. BINGKAI ANALISIS

Proses seleksi literatur yang telah dilakukan, dilakukan pemetaan berdasarkan abstrak, penulis, metode, judul dan kesimpulan. Berdasarkan hasil pencarian dari berbagai sumber diperoleh 10 literatur yang akan dijabarkan pada Tabel 1. Pemilahan artikel ini didasarkan pada jurnal yang telah terbit pada kurun waktu 6 tahun kebelakang dengan fokus utama yaitu, metode-metode yang dipakai sudah terbukti dapat mendeteksi kesalahan penulisan dari bahan uji coba yang mereka pakai seperti beberapa dokumen atau korpus dengan kumpulan kata dalam Bahasa Indonesia maupun Bahasa Inggris. Kemudian dari hasil tersebut di dapatkan nilai akurasi yang beragam sehingga metode yang dipakai sudah teruji dengan benar. Kebanyakan dari penelitian terdahulu mengacu pada perbandingan 2 metode saja, dengan hasil yang di dapat dari salah satu metode itu lebih baik atau lebih buruk.

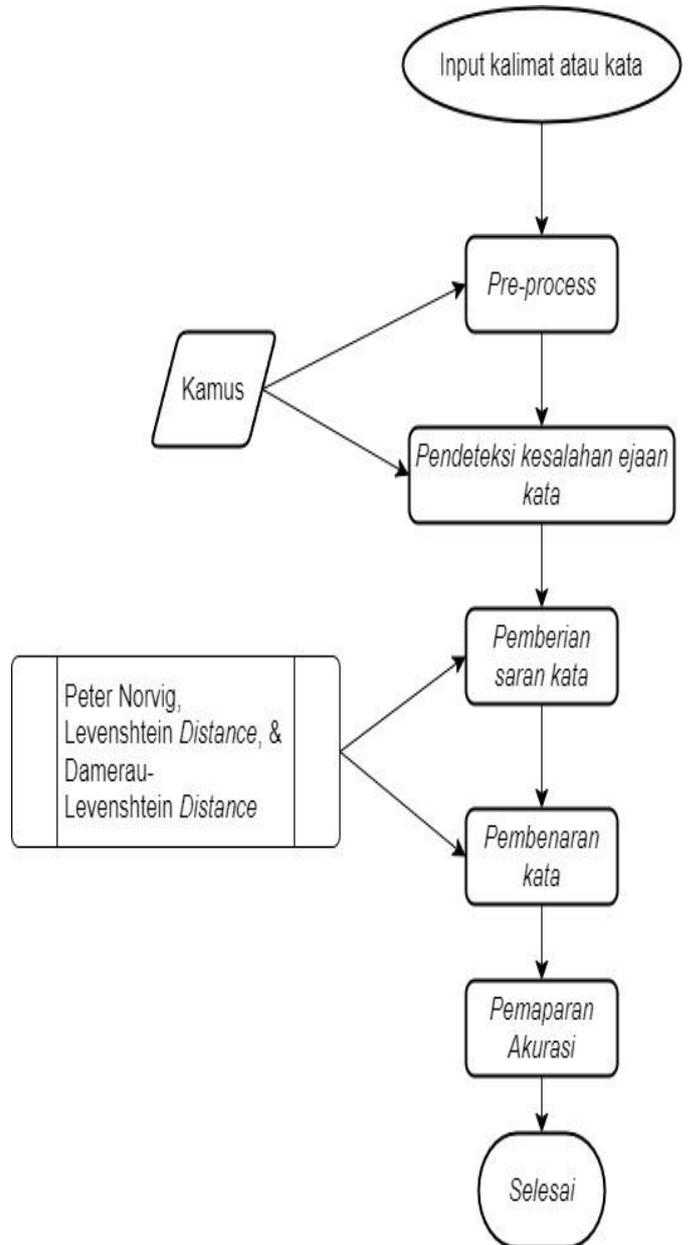
Tabel 1 Pemaparan Literatur

Literatur	Judul	Penulis	Metode	Hasil
[1]	Identifikasi Kesalahan Penulisan Kata (<i>Typographical Error</i>) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan <i>Levenshtein Distance</i>	Imam Cholissodin dan Rizal Setya Perdana	Metode N-Gram dengan <i>Levenshtein Distance</i>	Pengujian menyimpulkan bahwa metode <i>Levenshtein Distance</i> dapat mengidentifikasi kandidat kata yang sesuai dengan nilai yang diharapkan oleh pengguna. Presisi terbaik yang dihasilkan oleh sistem adalah 0,97 pada skenario pengujian jenis kesalahan tipografi "insertion", sedangkan recall terbaik yang dihasilkan oleh sistem adalah 1 pada skenario pengujian jenis kesalahan tipografi "substitution".
[3]	<i>Comparison of Levenshtein Distance Algorithm and Needleman-Wunsch Distance Algorithm for String Matching</i>	Khin Moe Myint Aung B.C.Sc.	<i>Levenshtein Distance</i> dengan <i>Needleman-Wunsch Distance</i>	Hasil dari eksperimen ini, Algoritma Jarak <i>Levenshtein</i> memiliki akurasi yang lebih baik daripada Algoritma Jarak <i>Needleman Wunsch</i> .
[4]	<i>Spelling Checker</i>	Nur Hamida,	Damerau	Eksperimen ini menunjukkan bahwa

	<i>using Algorithm Damerau Levenshtein Distance and Cosine Similarity</i>	Novi Yuliani dan Desty Rodiah	<i>Levenshtein Distance</i> dan <i>Cosine Similarity</i>	Damerau <i>Levenshtein Distance</i> dapat menghasilkan kandidat kata yang sesuai dengan empat jenis kesalahan kata, yaitu penambahan, penghapusan, penggantian, dan pergeseran. Selain itu, Algoritma Kesamaan Kosinus dapat menghitung nilai kemiripan antara kata yang salah dengan kandidat kata.
[5]	<i>Spelling Corrector Bahasa Indonesia dengan Kombinasi Metode Peter Norvig dan N-Gram</i>	Maya Simanjuntak, Novi Safriadi, Herry Sujaini	Peter Norvig dan N-Gram	Dari hasil eksperimen ini, penggabungan metode Peter Norvig <i>Spelling Corrector</i> dan N-Gram pada kalimat Bahasa Indonesia menghasilkan tingkat keakuratan sebesar 65,926% dalam menemukan satu saran kata. Keberhasilan penggabungan kedua metode ini sangat tergantung pada korpus yang digunakan. Semakin baik korpus yang digunakan, semakin baik pula sistem dalam menemukan kata-kata yang relevan.
[6]	<i>Fast and Accurate Spelling Correction Using Trie and Damerau-Levenshtein Distance Bigram</i>	Viny Christanti M, Rudy, dan Dali S. Naga	Trie dan Damerau-Levenshtein	Sistem yang diusulkan menghasilkan hasil terbaik berdasarkan penelitian, menggunakan bigram dan jarak Damerau-Levenshtein. Dalam hal ini, sistem mencapai akurasi tingkat kalimat sebesar 50%, akurasi tingkat kata sebesar 84,62%, dan waktu pemrosesan rata-rata per kalimat sekitar 18,89 ms.
[7]	<i>Damerau Levenshtein distance for Indonesian spelling correction</i>	Puji Santoso, Pundhi Yuliatwati, Ridwan Shalahuddin, dan Aji Prasetya Wibawa	<i>Levenshtein Distance</i> & Damerau-Levenshtein	Penerapan algoritma Damerau-Levenshtein untuk memperbaiki ejaan dalam cerita dongeng anak-anak menghasilkan hasil yang lebih baik daripada algoritma jarak <i>Levenshtein</i> . Tingkat akurasi pengujian menggunakan algoritma jarak <i>Levenshtein</i> sebesar 73%, sedangkan dengan algoritma Damerau-Levenshtein sebesar 75%. Selain itu, algoritma Damerau-Levenshtein memberikan saran kata yang lebih baik daripada algoritma jarak <i>Levenshtein</i> .

[8]	Penggunaan <i>Spelling Correction</i> dengan metode Peter Norvig dan N-gram	Ricky Martin, Dali Santun Naga, dan Viny Christanti Mawardi	Peter Norvig dan N-Gram	Berdasarkan pengujian program <i>Spelling Correction</i> menggunakan metode Peter Norvig dan n-gram, ditemukan bahwa akurasi hasil <i>Spelling Correction</i> mencapai 69,09% dengan menggunakan 55 dokumen sebagai data uji. Namun, jika program ini menggunakan lebih banyak data uji, kemungkinan akurasi hasilnya akan lebih rendah daripada 69,09%.
[9]	<i>Autocomplete and Spell Checking Levenshtein Distance Algorithm to Getting Text Suggest Error Data Searching in Library</i>	Muhammad Maulana Yulianto, Riza Arifudin, dan Alamsyah	Levenshtein Distance	Hasil proses pengimplementasian, Dalam proses pencarian data, menggunakan algoritma jarak Levenshtein yang memiliki tiga operasi pencocokan string, yaitu penghapusan, penambahan, dan penggantian. menggunakan metode ini di dapat nilai akurasi pada sistem sebesar 86% dengan 1055 data
[10]	Damerau Levenshtein Distance dengan Metode Empiris untuk Koreksi Ejaan Bahasa Indonesia	Aji Prasetya Wibawa, Pundhi Yuliawati, Puji Santoso, Ridwan Shalahuddin, dan I Made Wirawan	Damerau Levenshtein Distance dengan Metode Empiris	Penggunaan algoritma Damerau-Levenshtein Distance dengan menggabungkan metode empiris mencapai 97% akurasi. Menggunakan metode ini memiliki beberapa kelemahan seperti hanya bisa menghasilkan satu saran kata disetiap kesalahan kalimatnya.
[11]	<i>Identifying The Common Type of Spelling Error by Leveraging Levenshtein Distance and N-gram</i>	Margareta Hardiyanti	Levenshtein Distance & N-gram	Metode Levenshtein Distance memiliki <i>recall value</i> yang lebih tinggi dibanding metode N-gram, untuk memperbaiki kesalahan huruf pada kalimat dan membenarkan dua kata yang bersamaan dalam satu kalimat

Literatur yang sudah diseleksi kemudian di pilah akan berfokus pada 3 metode saja berdasarkan judul kajian literatur ini yaitu, Peter Norvig, Levenshtein *Distance*, dan Damerau-levenshtein *Distance*. Pada *spell checking* terdapat 4 komponen penting, *Pre-process*, pendeteksian kesalahan ejaan kata, pemberian saran kata sesuai kesalahan ejaan, pembenaran kata, dan pemaparan akurasi dari setiap metode.



Gambar 1. Contoh Penerapan Proses Perbandingan Metode
A. *Pre-Process*

Setelah penjabaran yang ditunjukkan pada Tabel 1, Pada literatur [1], [4], [7], dan [10] metode yang digunakan memiliki kombinasi dengan menggunakan metode N-Gram, yang dimana N-gram sendiri merupakan metode yang independen atau berdiri sendiri. Dalam pengembangan sistem *spell checking* itu sendiri terdapat berbagai macam metode yang bisa digunakan, namun pada kajian literatur ini hanya akan fokus ke beberapa metode saja seperti “Peter Norvig”, “Levenshtein *Distance*”, dan “Damerau Levenshtein *Distance*”.

Tahap ini merupakan awal mulainya tahapan pada suatu proses *spell checking* sesuai pada literatur ke [3] disebutkan *pre-process* merupakan proses mengatur sebuah data untuk mempersiapkan teks ke tahap selanjutnya. Persiapan teks ke tahap selanjutnya merupakan tahap dimana *Pre-Process* menghapus tanda bada dan angka dalam sebuah kalimat *Case folding* merupakan proses untuk menyamakan abjad di dalam suatu dokumen, dari huruf kapital menjadi huruf kecil, selain daripada 26 abjad akan diabaikan dan akan dianggap sebagai pemisah. Kemudian pada *pre-process*

terdapat tokenisasi atau proses untuk memecah suatu kalimat menjadi kalimat terkecil[4].

B. Pendeteksi Kesalahan Ejaan Kata

Pada pemrosesan ini akan dicari segala bentuk kesalahan yang ada pada suatu dokumen baik itu berupa kata atau berupa kalimat, yang dimana nantinya akan diberi peringatan ke pemilik dokumen tentang apa saja kesalahan yang terdapat pada dokumen tersebut. Mengacu pada literatur yang sudah dipaparkan terdapat dua tipe pendeteksi kesalahan ejaan yaitu, yang pertama kesalahan yang bersifat *non-word*. *Non-word* merupakan proses pengecekan kesalahan pada dokumen untuk mengecek huruf atau ejaan yang berlebih sehingga mengakibatkan *infinite list*. Sedangkan kesalahan yang kedua adalah *real-word* yaitu, kesalahan yang terdapat ambigu, dan struktur tata bahasa di setiap kalimat, serta beberapa kata yang tidak ada di dalam kamus atau database sistem disebut *Out of Vocabulary*[1].

C. Pemberian Saran Kata & Pembetulan Kata

Merujuk pada literatur ke [2]Pemberian saran kata metode baik Peter Norvig, Levenshtein *Distance*, dan Damerau Levenshtein *Distance*. Ke tiga metode tersebut akan melakukan probabilitas untuk mendeteksi kemungkinan terdekat diantara kalimat ejaan yang salah dan kalimat yang sudah tersedia di dalam data yang sudah disiapkan. Untuk contoh sederhananya metode akan melakukan seleksi calon yang akan menjadi kandidat kata benar melalui operasi *splits*, *deletion*, *transposition*, *substitution*, dan *insertion*. Untuk kalimat contoh akan menghasilkan: “ak”, “ku”, “auk”, “Ak”, “aKKu” dan “uKa”, kalimat akan digabung dari a sampai z. Ketika kandidat kata sudah melakukan perhitungan kemungkinan di dalam korpus, kata yang paling mendekati dari “Aku” adalah “aKKu”. Proses ini akan terus berlanjut untuk memproses kata terdekat yang bisa ditemukan[12].

D. Pemaparan Akurasi

Akurasi sendiri di dalam *spell checking* tidak selamanya bisa mencapai angka 100%. Contohnya saja dalam penggunaan metode Levenshtein *Distance* hanya menghasilkan 73% akurasi dan Damerau Laevenshtein *Distance* hanya 75% akurasi menurut literatur [7]. Lavenshtein *Distance* hanya bisa membenarkan 73 kesalahan dari 100 total kesalahan dan 27 kesalahan kata tidak dapat diperbaiki sedangkan Damerau-Levenshtein *Distance* hanya bisa membenarkan 75 kesalahan ejaan kata dari 100 total secara keseluruhan dan tidak dapat membenarkan 25 kesalahan tulis. Damerau-Levenshtein *Distance* mendapati peningkatan sebesar 2%. Kalimat-kalimat yang tidak dapat dibenarkan kesalahan penulisannya karena terlalu banyak pemberian saran kata dan dua kata yang tersambung atau tidak ada spasi diantara kalimat.

Contoh berikutnya pada metode Peter Norvig hanya menghasilkan 69.09% dengan memakai 55 dokumen untuk bahan uji dari literatur [13], hasil akurasi tersebut akan semakin menurun apabila data uji menggunakan data yang lebih banyak, tidak ditemukannya sebuah batasan sehingga tidak dapat mencakup kata umum di dalam kamus. Metode Peter Norvig sendiri sangat bergantung pada kualitas kamus atau korpus yang dijadikan sebagai dasar uji setiap kata.

Metode Peter Norvig sendiri tidak dapat melakukan pengoreksian terhadap kesalahan pada 2 huruf dalam 1 kata,

termasuk kalimat yang mengandung nama orang akan termasuk dalam kesalahan ejaan karena nama orang tidak termasuk daftar kata yang ada di KBBI.

IV. HASIL DAN PEMBAHASAN

Dari berbagai literatur yang dipilih, terdapat beberapa metode yang digunakan untuk klasifikasi dan pencarian. Secara umum, literatur yang diteliti menggunakan beberapa metode yang menjadi fokus pada penelitian ini ditambah dengan metode lainnya.

Pada penelitian [1], dan [11] melakukan perbandingan metode N-gram dan Levenshtein Distance, tetapi pada penelitian [9] hanya melakukan penelitian Levenshtein Distance namun pengertian dan hasil penelitian tersebut tidak jauh beda dengan 2 penelitian dengan membandingkan metode lainnya. Perbandingan metode N-gram dan Levenshtein Distance dibagi menjadi 5 jenis pengujian *Typographical Error* yaitu, berdasarkan jumlah dokumen, berdasarkan seluruh jenis jumlah kata kesalahan tipografi di dalam satu dokumen, jenis kesalahan tipografi *insertion* menurut jumlah kata dalam satu dokumen, pengujian kesalahan tipografi *deletion* berdasarkan jumlah kata di dalam satu dokumen, pengujian berdasarkan kesalahan tipografi *substitution* dalam satu dokumen. Untuk pengujian dengan kesalahan tipografi *insertion* merupakan jenis kesalahan tipografi di mana sebuah huruf ditambahkan ke dalam kata sehingga kata tersebut tidak terdaftar dalam kamus yang digunakan, *deletion* kata mengalami penghilangan huruf sehingga kata tersebut tidak dikenali dalam kamus yang digunakan, dan *substitution* terjadi ketika sebuah kata mengalami pergantian huruf dari kata yang seharusnya, sehingga kata tersebut tidak dikenali dalam kamus.

Seperti pada penelitian sebelumnya yang membahas tentang Levenshtein Distance, pada penelitian [3] sistem bereksperimen dengan memanfaatkan informasi lagu untuk kasus 1 dan kasus 2. Pada Kasus 1, digunakan 100 data untuk fitur Artis dalam kumpulan data lagu billboard dengan masukan pengguna yang sama dengan data set. Ementara itu, pada Kasus 2, digunakan 100 data untuk fitur judul lagu dalam kumpulan data lagu billboard dengan kesalahan ejaan 2 metode tersebut dibandingkan menggunakan f-score dan waktu eksekusi. Perbandingan dengan f-score rata-rata algoritma Levenshtein Distance memiliki tingkat akurasi yang lebih tinggi sedangkan untuk perbandingan menggunakan waktu eksekusi Needleman-Wunshc memiliki kompleksitas waktu yang lebih baik.

Sedangkan pada penelitian [4] dan [10] sedikit berbeda karena membahas tentang metode Damerau Levenshtein yaitu perkembangan dari metode Levenshtein *Distance*. Pada 2 penelitian ini terdapat kemiripan dengan membahas Damerau Levenshtein distance tetapi untuk perbandingan metode tersebut terdapat pembeda yaitu melakukan perbandingan dengan *Cosine Similarity* dan Metode Empiris. Penggunaan metode Damerau Levenshtein digunakan untuk menghitung jumlah operasi minimal dari satu string ke string lainnya dengan penambahan operasi transposisi, sedangkan *Cosine Similarity* fungsi untuk mengukur tingkat kemiripan antara dua vektor. Sedangkan metode Empiris melakukan pengoreksian kata dengan memisahkan string menjadi beberapa kata yang memiliki

kemungkinan terbesar kemudian mencocokkannya dengan *database*.

Dalam penelitian [5] dan [8] perbandingan metode sama-sama menggunakan Peter Norvig dan N-gram. Sistem Peter Norvig yang diuji pada penelitian ini tidak dapat memeriksa 2 huruf yang terdapat kesalahan di dalam 1 kata. Kombinasi yang digunakan antara Peter Norvig dan N-gram memiliki 65.926% keberhasilan tepat sasaran untuk menemukan saran pada satu kata.

V. KESIMPULAN

Berdasarkan dari hasil studi literatur, perbandingan metode *spell checking* yaitu, Peter Norvig, Levenshtein Distance, dan Damerau Levenshtein Distance. Tidak dapat menghasilkan ketepatan yang sempurna dalam pengecekan dan membenaran kesalahan tetapi dengan menggunakan 3 metode tersebut dapat membuat sistem *spell checking* menjadi lebih beragam dalam pendeteksian kesalahan penulisan dalam dokumen. Sistem dengan menggunakan 3 metode tersebut dapat diimplementasikan juga ke dalam website yang terhubung dengan server sehingga masyarakat dapat melakukan pengecekan dokumen secara pribadi di *device* masing-masing. Diharapkan penelitian yang akan datang dengan topik yang mirip atau bahkan sama dapat mengembangkan dan menambahkan metode yang akan dipakai di *spell checking*.

DAFTAR PUSTAKA

- [1] I. Cholissodin and R. Setya Perdana, "Identifikasi Kesalahan Penulisan Kata (Typographical Error) pada Dokumen Berbahasa Indonesia Menggunakan Metode N-gram dan Levenshtein Distance Artificial Neural Network Applications View project Twitter Sentiment Analysis View project," 2017. [Online]. Available: <https://www.researchgate.net/publication/32336572>
- [2] T. M. Fahrudin, I. Sa'diyah, Latipah, I. Z. A. Illah, C. C. B. Lirna, and B. S. Acarya, "A Rule-based Spelling Checker for Correcting Punctuation Errors in Indonesia Text using KEBI 1.0 Checker," *Galaxy Science*, May 2022. doi: 10.11594/nstp.2022.2433.
- [3] K. M. M. B. C. sc Aung, "COMPARISON OF LEVENSHTAIN DISTANCE ALGORITHM AND NEEDLEMAN-WUNSCH DISTANCE ALGORITHM FOR STRING MATCHING KHIN MOE MYINT AUNG COMPARISON OF LEVENSHTAIN DISTANCE ALGORITHM AND NEEDLEMAN-WUNSCH DISTANCE ALGORITHM FOR STRING MATCHING," 2019.
- [4] N. Hamidah, N. Yusliani, and D. Rodiah, "Spelling Checker using Algorithm Damerau Levenshtein Distance and Cosine Similarity," 2020. [Online]. Available: <http://sjia.ejournal.unsri.ac.id>
- [5] M. S. Simanjuntak, H. Sujaini, and N. Safriadi, "Spelling Corrector Bahasa Indonesia dengan Kombinasi Metode Peter Norvig dan N-Gram," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 1, p. 17, Jun. 2018, doi: 10.26418/jp.v4i1.24075.
- [6] M. V. Christanti, Rudy, and D. S. Naga, "Fast and accurate spelling correction using trie and Damerau-Levenshtein distance bigram," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 16, no. 2, pp. 827–833, Apr. 2018, doi: 10.12928/TELKOMNIKA.v16i2.6890.
- [7] P. Santoso, P. Yuliawati, R. Shalahuddin, and A. P. Wibawa, "Damerau Levenshtein Distance for Indonesian Spelling Correction," *Jurnal Informatika*, vol. 13, no. 2, p. 11, Jul. 2019, doi: 10.26555/jifo.v13i2.a15698.
- [8] R. Martin, D. S. Naga, and V. C. Mawardi, "PENGUNAAN SPELLING CORRECTION DENGAN METODE PETER NORVIG DAN N-GRAM."
- [9] M. M. Yulianto, R. Arifudin, and Alamsyah, "Autocomplete and Spell Checking Levenshtein Distance Algorithm to Getting Text Suggest Error Data Searching in Library," *Scientific Journal of Informatics*, vol. 5, no. 1, pp. 2407–7658, 2018, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>
- [10] A. P. Wibawa, P. Yuliawati, P. Santoso, R. Shalahuddin, and I. M. Wirawan, "Damerau Levenshtein Distance dengan Metode Empiris untuk Koreksi Ejaan Bahasa Indonesia," *ILKOM Jurnal Ilmiah*, vol. 12, no. 3, pp. 176–182, Dec. 2020, doi: 10.33096/ilkom.v12i3.600.176-182.
- [11] M. Hardiyanti, "Identifying The Common Type of Spelling Error by Leveraging Levenshtein Distance and N-gram," *Scientific Journal of Informatics*, vol. 8, no. 1, 2021, doi: 10.15294/sji.v8i1.xxxxx.
- [12] T. M. Fahrudin, I. Sa'diyah, L. Latipah, I. Z. Atha Illah, C. C. Bey Lirna, and B. S. Acarya, "KEBI 1.0: Indonesian Spelling Error Detection System for Scientific Papers using Dictionary Lookup and Peter Norvig Spelling Corrector," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 12, no. 2, p. 78, Aug. 2021, doi: 10.24843/lkjiti.2021.v12.i02.p02.
- [13] R. Martin, D. Santun Naga, and V. Christanti Mawardi, "PENGUNAAN SPELLING CORRECTION DENGAN METODE PETER NORVIG DAN N-GRAM."