

Aplikasi Pendeteksi Kesalahan Ejaan Bahasa Indonesia pada Karya Ilmiah Bidang Ilmu Komputer Menggunakan KEBI 1.0 Checker

Tresna Maulana Fahrudin¹, Ilmatius Sa'diyah², Latifah³,

Ibnu Zahy' Atha Illah⁴, Cagiva Chaedar Bey Lirna⁵, Burhan Syarif Acarya⁶

^{1,2,4,5,6} Program Studi Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur
{¹tresna.maulana.ds,²ilmatus.sisfo}@upnjatim.ac.id,

{⁴20083010016,⁵20083010020,⁶20083010004}@student.upnjatim.ac.id

³ Program Studi Teknik Informatika, Universitas Narotama

[³latifah.rifani@narotama.ac.id](mailto:latifah.rifani@narotama.ac.id)

*Corresponding author email: tresna.maulana.ds@upnjatim.ac.id

Abstrak— Aktivitas menulis menjadi kewajiban bagi akademisi terutama di lingkungan perguruan tinggi yang erat kaitannya dengan publikasi karya ilmiah. Luaran karya ilmiah ini antara lain berupa skripsi, tesis, disertasi, buku, artikel, modul, tutorial, dan lainnya. Namun, aktivitas menulis ini tidak diikuti dengan pemahaman ejaan bahasa Indonesia dengan baik dan benar sehingga masih banyak ditemui kesalahan ejaan dalam penulisan artikel berbahasa Indonesia. Latar belakang penulis juga memengaruhi pemilihan kosakata yang dituangkan dalam karya ilmiah, misalnya di bidang ilmu komputer. Karya ilmiah berbahasa Indonesia di bidang ilmu komputer cukup banyak ditemui penggunaan kombinasi kata dalam bahasa Indonesia maupun kata dalam bahasa Inggris yang digunakan untuk menunjukkan suatu nama perangkat, algoritma, metode dan istilah lainnya. Oleh karena itu, penelitian ini berinisiatif untuk mengembangkan aplikasi berbasis web untuk mendeteksi kesalahan ejaan kata tidak baku dan kesalahan pengetikan kata (*typo*) yang disebut KEBI 1.0 Checker. Kehadiran KEBI untuk memudahkan semua penulis karya ilmiah dalam menghasilkan karya tulis ilmiah yang berkualitas berdasarkan Pedoman Umum Ejaan Bahasa Indonesia yang diterbitkan oleh Badan Pengembangan dan Pengembangan Bahasa. Hasil eksperimen menunjukkan bahwa aplikasi telah memberikan kinerja akurasi terbaik untuk mengoreksi kata-kata tidak baku dan kesalahan ketik masing-masing mencapai 100% dan 49,10%. Namun, KEBI belum mampu menangani 681 kata yang tidak bermakna dan 327 istilah ilmu komputer yang terdapat pada artikel ilmiah. Waktu pemrosesan aplikasi cukup rendah, rata-rata waktu yang dibutuhkan untuk mengoreksi kata tidak baku adalah 0,24 detik dan kata yang salah ketik adalah 21,72 detik. KEBI 1.0 Checker cukup berguna bagi akademisi, tetapi perlu meningkatkan kosakata korpus besar di berbagai bidang ilmu untuk mengoreksi kata-kata *typo*.

Kata Kunci— Kesalahan Ejaan, Bahasa Indonesia, Karya Ilmiah, Ilmu Komputer, KEBI 1.0 Checker

I. PENDAHULUAN

Aktivitas menulis telah menjadi kewajiban bagi civitas akademik, baik dosen maupun mahasiswa, dalam menunjang keilmuan secara akademik sekaligus sebagai tuntutan tridarma perguruan tinggi. Dosen dituntut untuk terus menulis artikel ilmiah hasil penelitian dan pengabdian kepada masyarakat tiap

semester. Begitu pula dengan mahasiswa yang memiliki kewajiban menulis untuk memenuhi tugas dari dosen dan sebagai syarat kelulusan dari kampus, yaitu menulis skripsi. Sayangnya, aktivitas menulis ini tidak diikuti dengan pemahaman ejaan bahasa Indonesia yang baik dan benar sehingga dosen dan mahasiswa masih banyak melakukan kesalahan ejaan dalam penulisan artikel berbahasa Indonesia [1].

Kesalahan tersebut meliputi seluruh unsur kebahasaan, yaitu penulisan tanda baca, penulisan huruf, penulisan kata, penulisan kata yang berasal dari bahasa asing/daerah (kata serapan), penggunaan kata berimbuhan, penulisan angka dan lambang bilangan bahkan penulisan kalimat yang tidak efektif. Bahkan, saat artikel ilmiah itu sudah terbit sebagai referensi keilmuan di masyarakat, baik dalam bentuk artikel di jurnal ilmiah atau prosiding, kesalahan ejaan masih ada. Misalnya, pada artikel [2] terdapat kalimat "Literatur terpilih yang akan dibahas sebanyak 15 artikel". Kalimat itu masih mengandung kesalahan ejaan pada penulisan angka dan lambang bilangan. Angka 15 seharusnya ditulis menggunakan huruf karena masih terdiri atas dua kata. Fakta tersebut menunjukkan bahwa kualitas tulisan Indonesia di tingkat perguruan tinggi sangat memprihatinkan, padahal ragam bahasa Indonesia baku digunakan oleh orang-orang terpelajar dalam pendidikan formal dan juga dipelajari dari pendidikan formal [3].

Kesalahan-kesalahan yang dilakukan dosen kemudian menjadi siklus di lingkungan akademik karena dosen perlu memberikan bimbingan dalam penulisan skripsi, tesis, dan disertasi kepada mahasiswa. Terakhir, pelajar juga melakukan kesalahan ejaan yang sama dalam bahasa Indonesia. Murtiningsih [4] menyatakan bahwa sebanyak 69,2% pelajar melakukan kesalahan dalam menulis kata-kata yang tidak sesuai konteks penulisan dan 17,4% pelajar melakukan kesalahan dalam menulis imbuhan dan preposisi. Kesalahan ini disebut kesalahan ketik [5] dan sering dilakukan oleh mahasiswa saat menulis makalah atau tugas skripsi.

Kesalahan ejaan bahasa Indonesia yang terjadi dapat berupa kesalahan yang disengaja maupun tidak disengaja [6]. Menurutnya, ada beberapa faktor, yaitu ketidaktepatan pengetikan, kurangnya pengetahuan tentang bahasa baku

penulisan dan tanda baca yang benar, kebiasaan menggunakan bahasa nonformal yang digunakan sehari-hari sehingga muncul dalam karya tulis ilmiah, dan ketidaktahuan prinsip-prinsip dari ejaan bahasa Indonesia. Selama ini, tulisan-tulisan ilmiah dalam bahasa Inggris telah terbantu dalam perbaikan ejaannya dengan adanya aplikasi *grammarly checker*. Dari aplikasi tersebut, kesalahan ejaan yang terjadi langsung ditandai, bahkan disarankan dengan kata pengganti yang tepat [7].

Karya ilmiah berbahasa Indonesia di bidang ilmu komputer cukup banyak ditemui penggunaan kombinasi kata dalam bahasa Indonesia maupun kata dalam bahasa Inggris yang digunakan untuk menunjukkan suatu nama perangkat, algoritma, metode dan istilah lainnya. Terkadang penulis yang berlatar belakang ilmu komputer juga perlu memperhatikan penggunaan kata baku dan ketepatan dalam mengetik untuk merepresentasikan suatu kalimat. Ini penting agar penulis dari bidang keilmuan tetap mampu mempertahankan kualitas karya tulis ilmiah sesuai dengan kaidah dalam bahasa Indonesia.

Oleh karena itu, penelitian ini berinisiatif untuk mengembangkan aplikasi berbasis web untuk mendeteksi kesalahan ejaan kata tidak baku dan kesalahan pengetikan kata (*typographical error*) yang disebut KEBI 1.0 Checker. KEBI adalah kependekan dari Kesalahan Ejaan Bahasa Indonesia. Kehadiran KEBI untuk memudahkan semua penulis karya ilmiah dalam menghasilkan karya tulis ilmiah yang berkualitas berdasarkan Pedoman Umum Ejaan Bahasa Indonesia yang diterbitkan oleh Badan Pengembangan dan Pengembangan Bahasa. Dosen dan mahasiswa bidang ilmu komputer dapat memanfaatkan aplikasi ini untuk aktivitas *self-editing* setelah menulis karya ilmiah sebelum dikirim ke penerbit jurnal atau seminar atau konferensi. KEBI juga diharapkan berdampak positif pada proses penyuntingan agar menjadi lebih ringan [8]. Pada artikel ini, aplikasi pendeteksi kesalahan ejaan bahasa Indonesia pada karya ilmiah bidang ilmu komputer menggunakan KEBI 1.0 Checker diuraikan lebih detail.

II. PENELITIAN TERKAIT

Pembangunan aplikasi pendeteksi dan pengoreksi kesalahan ejaan sudah pernah dilakukan beberapa peneliti. Yessy Prima Putri dan Ridwan Lawson membangun aplikasi pengkoreksi kesalahan ejaan dan padanan kata pada tugas akhir mahasiswa menggunakan metode *Full Text Indexing* yang merupakan metode untuk mencari kesalahan dalam sebuah teks berbasis web. Hasil dari penelitian ini adalah berupa aplikasi pendeteksi kesalahan penulisan naskah dokumen skripsi yang di dalamnya terdapat fungsi pengecekan kesalahan penulisan dan penyimpanan daftar pustaka dan daftar gambar. Aplikasi ini juga menyediakan tempat untuk menambahkan kosakata yang baru untuk pengembangan kata yang tersedia pada aplikasi [9].

Junar Frando dkk juga membangun aplikasi pengoreksi kesalahan penulisan bahasa Indonesia dengan menggunakan metode *Jaro Winkler Distance* dengan membandingkan kesamaan karakter penyusun dari dua buah string [10]. Pemindai kata-kata pada dokumen yang tidak terdaftar dalam kamus suatu bahasa tertentu merupakan cara dalam aplikasi ini. Pengumpulan data pada basis data berdasarkan pada

Kamus Besar Bahasa Indonesia (KBBI). Hasil dari penelitian ini adalah diperlukan waktu 25,5 detik untuk menguji jumlah total 37.501 kata pada 3 sampel dokumen dengan total persentase akurasi sebesar 77,23% dalam memberikan saran perbaikan. Namun, ada beberapa faktor yang menyebabkan aplikasi mengalami kegagalan dalam memberikan saran perbaikan, yaitu apabila ada kata yang ditulis berdempetan tanpa spasi dan kurangnya kata-kata dalam basis data aplikasi ini.

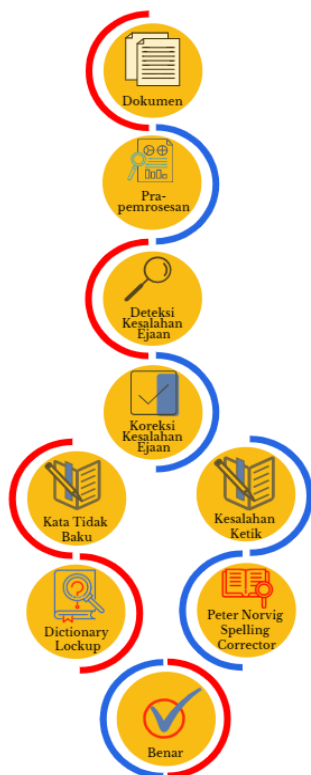
Pada penelitian yang lain, pembangunan aplikasi pendeteksi kesalahan ejaan juga dilakukan oleh Tusty Nadia Maghfira dkk. Mereka menggunakan metode *Dictionary Lookup* dan *Damerau-Levenshtein Distance* yang merupakan pengembangan algoritma *Levenshtein Distance* untuk menentukan suatu ejaan kata bernilai benar atau salah berdasarkan *Lexical Resource* [11]. Penelitian ini menguji 5 dokumen jurnal dengan mengecualikan persamaan dan daftar pustaka dari dokumen-dokumen tersebut. Dari penelitian ini didapatkan bahwa metode *Dictionary Lookup* dan *Damerau-Levenshtein Distance* lebih baik daripada metode *Levenshtein Distance* dengan nilai presisi sebesar 78% dan *recall* sebesar 100% yang didapatkan pada skenario pengujian kedua. Selain itu, hasil pengujian menunjukkan bahwa tidak ada pengaruh yang signifikan antara jumlah kesalahan ejaan kata dan jumlah kata dalam dokumen terhadap kinerja koreksi sistem. Kelengkapan kata pada kamus sebagai acuan kandidat koreksi kata dan tipe kesalahan ejaan kata lah yang memiliki pengaruh lebih besar.

Dari uraian penelitian terkait di atas, aplikasi pendeteksi kesalahan ejaan telah diaplikasikan untuk memeriksa padanan kata pada tugas akhir, kesalahan penulisan dan rekomendasi kata yang tepat. Beberapa metode yang telah digunakan oleh peneliti tersebut antara lain metode *Full Text Indexing*, *Jaro Winkler Distance*, *Dictionary Lookup* dan *Damerau-Levenshtein Distance*. Namun, KEBI 1.0 Checker menawarkan fasilitas yang berbeda, KEBI dibangun dalam bentuk aplikasi berbasis web yang mampu mendeteksi dan mengoreksi kesalahan ejaan kata tidak baku menggunakan metode *Dictionary Lookup* serta kesalahan pengetikan kata menggunakan metode *Peter Norvig Spelling Corrector* dalam satu platform. Pada penelitian ini, KEBI diuji kemampuannya untuk mendeteksi kesalahan ejaan kata tidak baku dan kesalahan pengetikan kata pada karya ilmiah di bidang ilmu komputer.

III. DESAIN SISTEM

Rancangan sistem yang diusulkan dalam penelitian KEBI 1.0 diilustrasikan pada Gambar 1 yang terdiri dari empat proses yaitu prapemrosesan, deteksi kesalahan ejaan, koreksi kesalahan ejaan, dan hasil akhir mendapatkan koreksi kata yang benar. Proses diawali dengan menyiapkan dokumen karya ilmiah di bidang ilmu komputer yang akan diperiksa, kemudian dilakukan prapemrosesan yang terdiri dari tokenisasi dan mengubah tiap token menjadi karakter huruf kecil. Selanjutnya, merancang sistem dengan memiliki kemampuan mendeteksi dan mengoreksi kesalahan ejaan kata tidak baku dan kesalahan pengetikan kata berdasarkan penulisan bahasa Indonesia yang baik dan benar sesuai KBBI dan PUEBI. Karya ilmiah bidang ilmu komputer tersebut

nantinya akan dideteksi kata-kata yang tidak baku menggunakan *Dictionary Lookup* dan kesalahan pengetikan kata menggunakan *Peter Norvig Spelling Corrector*.



Gbr 1. Usulan perancangan desain sistem penelitian KEBI 1.0 Checker

A. Karya Ilmiah di Bidang Ilmu Komputer

Karya tulis ilmiah adalah suatu karya tulis yang disusun berdasarkan pendekatan dan metode ilmiah (aplikasi dari metode ilmiah) yang ditujukan untuk kelompok pembaca tertentu. Menyusun suatu karya tulis ilmiah bukanlah suatu pekerjaan yang mudah dan sederhana. Penulis harus menguasai keterampilan dan pengetahuan bahasa yang baik dan benar [12]. Indikator karya ilmiah dianggap berkualitas salah satunya jika mengikuti aturan penulisan bahasa Indonesia yang baik dan benar sesuai KBBI dan PUEBI.

Hasil pembelajaran dan eksperimen di bidang ilmu komputer dapat dituangkan ke dalam bentuk karya tulis ilmiah dan dapat dipertanggungjawabkan secara ilmiah. Ketika penulis membuat suatu karya tulis ilmiah misalnya pada bidang ilmu komputer sering dijumpai kesalahan penulisan bahasa Indonesia. Kesalahan tersebut dapat berupa kurangnya pengetahuan penulis akan bahasa Indonesia yang baik dan benar sesuai KBBI dan PUEBI, kelalaian penulis yang tidak disengaja, dan beberapa hal lain yang memungkinkan terjadinya kesalahan penulisan bahasa Indonesia. Selain itu, penggunaan kosakata pada bidang ilmu komputer juga cukup beragam, mulai dari penggunaan istilah untuk penyebutan suatu nama perangkat, metode, algoritma dan lainnya yang menyerap istilah kata dalam bahasa Inggris.

Kosakata dalam bidang ilmu komputer juga cukup beragam sesuai dengan bidang kajian yang dibahas, misalnya kosakata subbidang informatika, sistem informasi, teknologi informasi, teknik komputer dan sains data. Bahkan tiap subbidang juga memiliki klasifikasi konsentrasi keilmuan yang cukup beragam sehingga memengaruhi pemilihan kosakata yang diserap dari bahasa Inggris.

Berdasarkan Pedoman Umum Ejaan Bahasa Indonesia edisi keempat yang diterbitkan oleh Badan Pengembangan dan Pembinaan Bahasa Kementerian Pendidikan dan Kebudayaan pada tahun 2016, kesalahan ejaan bahasa Indonesia meliputi pemakaian huruf, penulisan kata, dan pemakaian tanda baca. Selain itu, kesalahan pengetikan kata juga termasuk ke dalam penyimpangan penulisan bahasa Indonesia yang baik dan benar sesuai KBBI dan PUEBI [13].

Pada penelitian ini hanya difokuskan pada deteksi kata tidak baku dan kesalahan pengetikan kata. Suatu kata dapat disebut sebagai kata tidak baku jika kata yang dipakai tidak sesuai dengan kaidah bahasa Indonesia. Kata baku merupakan sebuah kata yang digunakan sudah sesuai dengan pedoman atau kaidah bahasa yang sudah ditentukan [14], sedangkan kesalahan pengetikan kata adalah kesalahan yang dibuat oleh penulis pada saat proses mengetik yang berakibat pada penulisan kata yang dihasilkan tidak sesuai dengan kamus yang terdapat di KBBI. Apabila dalam pengetikan dokumen terdapat tulisan yang salah, maka dapat mengubah arti dan mengubah pemahaman bagi pembacanya [15].

B. Deteksi dan Koreksi Kesalahan Ejaan

Deteksi dan koreksi kesalahan ejaan merupakan bagian penting setelah dokumen melewati tahap prapemrosesan. Tahap prapemrosesan data merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, prapemrosesan data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem [16]. Prapemrosesan yang dilakukan pada penelitian ini adalah *tokenization* dan *case folding*. Karakter selain huruf alfabet dan gambar akan dihilangkan karena tidak diperlukan dalam deteksi dan koreksi kesalahan ejaan.

Setelah prapemrosesan selesai, dokumen dapat dilanjutkan pada proses deteksi kesalahan ejaan. Deteksi kesalahan ejaan merupakan tahap pemeriksaan kata dalam sebuah bahasa. Setiap kata akan diperiksa berdasarkan kamus yang telah terdaftar dalam sistem aplikasi KEBI 1.0 Checker. Setelah kata yang tidak sesuai dengan kamus ditemukan, proses selanjutnya yaitu melakukan koreksi kesalahan ejaan. Sistem akan menentukan bahwa ejaan dalam teks memiliki kesalahan, kemudian akan dilakukan pencarian dan membuat saran untuk memperbaiki kesalahan dari kandidat kata tersebut.

Koreksi kesalahan ejaan dapat dibedakan menjadi tiga jenis, yaitu:

- Koreksi otomatis: cara kerja sistem yang mengoreksi ejaan secara langsung.
- Saran koreksi: cara kerja sistem yang akan memberikan saran untuk perbaikan ejaan, tetapi menunggu tindakan dari penulis untuk memperbaikinya.

- Daftar saran: cara kerja sistem yang akan memberikan beberapa pilihan saran koreksi ejaan dan penulis diperbolehkan untuk memilih yang sesuai.

C. Dictionary Lookup

Seiring meningkatnya dokumen tertulis menuntut pula peningkatan kualitas dan keterampilan menulis masyarakat, salah satu kualitas dan keterampilan yang perlu dikuasai oleh masyarakat adalah penggunaan kata baku yang tepat. Namun, pada kenyataannya masyarakat masih sering melakukan kesalahan dalam penggunaan kata baku yang tepat dalam penulisan naskah dokumen tertulis.

Kesalahan penggunaan kata baku yang sering terjadi di masyarakat dalam menulis mendorong peneliti untuk menentukan solusi permasalahan dengan mengembangkan aplikasi KEBI yang di dalamnya terdapat salah satu fitur deteksi kata tidak baku. Metode *Dictionary Lookup* menjadi alternatif yang digunakan dalam membangun fitur tersebut. Metode *Dictionary Lookup* merupakan metode yang sering digunakan untuk menentukan kesalahan *non-word error*. Proses yang dilakukan pada metode ini yaitu dengan memeriksa ketersediaan kata tersebut pada kamus. Jika kata tersebut tidak tersedia, maka kata itu dianggap bukan kata (*non-word*). Metode ini termasuk cara yang efektif [11] yang ditawarkan KEBI dalam fitur deteksi kata tidak baku pada suatu karya tulis ilmiah.

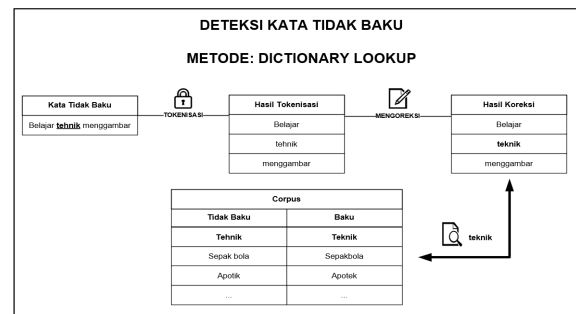
Menurut kaidah penulisan kata bahasa Indonesia, penggunaan kata baku dalam karya tulis ilmiah sangatlah penting. Kata baku biasanya digunakan dalam kalimat formal atau dalam berbagai bahasa standar, termasuk bahasa lisan dan tulisan. Kata baku bahasa Indonesia juga memiliki ciri-ciri sebagai berikut [17]:

- Baik lisan maupun tulisan, kata baku digunakan dalam acara resmi, seperti surat resmi, undang-undang, karya ilmiah, laporan penelitian, dan lainnya. Variasi bahasa yang baku tidak akan terganggu oleh dialek atau aksen tertentu.
- Baik lisan maupun tulisan, kata baku menggunakan istilah yang berlaku dalam Pedoman Ejaan Umum Bahasa Indonesia (PUEBI).
- Baik lisan maupun tulisan, varietas standar secara jelas dan lengkap menjalankan fungsi gramatikal subjek, predikat, dan objek.

Sebuah kata dapat disebut kata yang tidak baku jika kata yang digunakan tidak sesuai dengan aturan bahasa Indonesia. Ketidakbakuan sebuah kata tidak hanya disebabkan oleh kesalahan penulisan, tetapi juga dapat disebabkan oleh pengucapan yang salah dan persiapan kalimat yang salah. Kata tidak baku ini sering muncul dalam kehidupan sehari-hari kita [4]. Beberapa contoh kesalahan ejaan bahasa Indonesia dalam menulis kata tidak baku sebagai berikut:

- Kata tidak baku: *tehnik*, sedangkan kata baku: *teknik*.
- Kata tidak baku: *sepak bola*, sedangkan kata baku: *sepakbola*.

- Kata tidak baku: *apotik*, sedangkan kata baku: *apotek*.



Gbr 2. Desain sistem proses deteksi kesalahan ejaan kata tidak baku menggunakan metode *Dictionary Lookup*

Gambar 2 menunjukkan bagaimana proses mendeteksi kesalahan ejaan dalam kalimat yang berisi berbagai kata tidak baku bekerja, misalnya kalimat mengandung kata "tehnik" setelah tokenisasi. Kalimat yang berisi kata "tehnik" akan diperiksa dan dicocokkan dengan kata kunci yang berisi kumpulan pasangan kata-kata baku dan kata tidak baku dalam Indonesia di korpus. Kata kunci ini dicocokkan dengan input berupa kata tidak baku. Jika ditemukan kecocokan, maka kata baku dapat menggantikan kata tidak baku pada susunan kalimat sebelumnya. Hasilnya kata tidak baku "tehnik" akan diganti dengan kata baku "teknik". Proses ini adalah proses yang disebut *key:value* yang dapat diterapkan ke metode pencarian kamus.

D. Peter Norvig Spelling Corrector

Kesalahan ketik atau disebut *typo* sering dilakukan oleh penulis saat mengetik huruf pada dokumen digital. Semakin banyak halaman dokumen yang diketik, maka semakin besar kemungkinan kesalahan penulis dalam mengetik kata-kata dalam dokumen tersebut. Hal ini dapat dihindari jika penulis memeriksa halaman dokumen lebih awal sebelum melanjutkan ke halaman dokumen berikutnya. Contoh kasus kesalahan pengetikan kata dapat terjadi ketika siswa mengetik jawaban pertanyaan esai dan jawaban singkat. Dalam kasus lain, seseorang terkadang salah mengetik URL ketika mencoba mengakses sebuah situs web yang akhirnya mengarah ke halaman situs web yang salah. Seseorang juga terkadang membuat kesalahan ketika mengisi profil mereka di sebuah formulir seperti nama, tanggal lahir, alamat dan afiliasi.

Typo dapat dikategorikan menjadi dua jenis, yaitu kesalahan *non-word* dan kesalahan *real-word*. Kesalahan *non-word* adalah kesalahan yang tidak memiliki arti dalam kamus, sedangkan kesalahan *real-word* adalah kata yang tertulis benar atau memiliki arti dalam kamus, tetapi tidak dimaksudkan dalam kalimat dan memiliki arti yang berbeda. Penelitian ini berfokus pada kesalahan *non-word* karena seseorang sering melakukan kesalahan ini.

TABEL I

PROSES KOREKSI EJAAN *PETER NORVIG* UNTUK MENEMUKAN KOREKSI KATA MENGGUNAKAN *SPLITS, DELETION, TRANSPOSITION, SUBSTITUTION, DAN INSERTION*

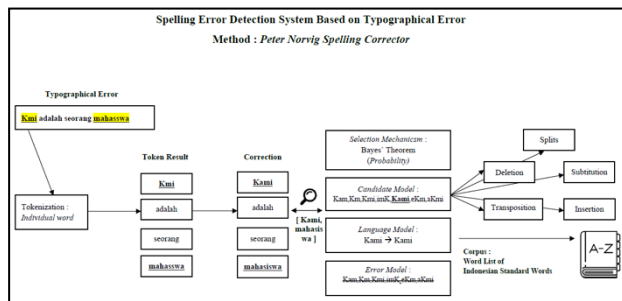
Incorrect words (typo)	'sbelum'						
Splits	", 'sbelum'	's', 'belum'	'sb', 'elum'	'sbe', 'lum'	sbel', 'um'	'sbelu', 'm'	'sbelum', "
Deletion	'belum'	'selum'	'sblum'	'sbeum'	'sbelm'	'sbelu'	
Transposition	'bselum'	'seblum'	'sbleum'	'sbeulm'	'sbelmu'		
Substitution	'abelum'	'saelum', 'sbelum'	'sbalum'	'sbeaum'	'sbelam'	'sbelua'	
	'bbelum'		'sbbelum'	'sbebum'	'sbelbm'	'sbelub'	
	
	
	'zbelum'	'szelum'	'sبزlum'	'sbezum'	'sbelzm'	'sbeluz'	
Insertion	'asbelum'	'sabelum'	'sbaelum'	'sbealum'	'sbelaum'	'sbeluam'	'sbeluma'
	'bsbelum'	'sbbelum'	'sbbelum'	'sbeblum'	'sbelbum'	'sbelubm'	'sbelumb'

	'esbelum'	'sebelum'	'sbeelum'	'sbeelum'	'sbeleum'	'sbeluem'	'sbelume'

	'zsbelum'	'szbelum'	'sبزelum'	'sbezlum'	'sbelzum'	'sbeluzm'	'sbelumz'
Correct words	'sebelum'						

Beberapa contoh *typo* pada kesalahan *non-word* sebagai berikut:

- Kata yang benar: "Kami", sedangkan *typo*: "Kmi".
- Kata yang benar: "mahasiswa", sedangkan *typo*: "mahasswa".
- Kata yang benar: "adalah", sedangkan *typo*: "adlah".

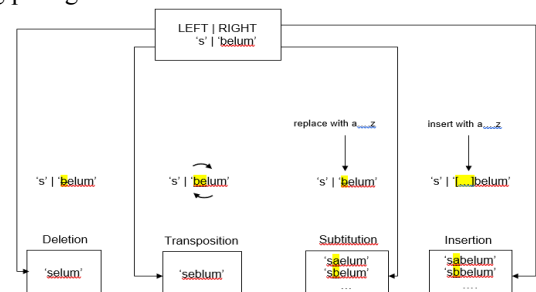


Gbr 3. Desain sistem proses deteksi kesalahan ejaan kata *typo* menggunakan metode *Peter Norvig Spelling Corrector*

Peter Norvig Spelling Corrector merupakan sebuah algoritma yang mampu mengubah kata yang mengandung kesalahan ejaan, lalu dicarikan kandidat-kandidat kata yang benar melalui proses pemisahan kata menjadi dua bagian dan sejumlah suntingan yang mengubah bentuk kata ke suatu bentuk ke bentuk yang lain. Pendekatan kata pada metode *Peter Norvig* dapat menghasilkan semua kemungkinan kata dengan semua operasi *edit-distance* yaitu operasi penyisipan (*insertion*), operasi substitusi (*substitution*), operasi penukaran (*transposition*), dan operasi penghapusan (*deletion*) dari kata yang terdeteksi *typo* dan mencarinya dalam kamus. Proses operasi diterapkan untuk semua huruf pada kata

typo secara bergantian. Setiap kali operasi bekerja akan menghasilkan satu kata yang berbeda dari kata awal, kemudian kata-kata tersebut diperiksa dan dicocokkan di dalam korpus. Selanjutnya, kandidat kata yang berhasil dinilai sesuai dengan daftar kata di korpus akan dijadikan pengganti kata *typo*.

Gambar 3 menunjukkan desain sistem proses deteksi kesalahan ejaan kata *typo* menggunakan *Peter Norvig Spelling Corrector*. Metode ini menggunakan konsep probabilitas untuk memprediksi kemungkinan kedekatan antara kata *typo* dengan kata-kata yang tersedia dalam korpus. Sebagai contoh, terdapat kata *typo* "Kmu" (kata yang benar: "Kamu"). Metode ini akan mencari kandidat kata yang mendekati kata sebenarnya dengan menggunakan model kandidat seperti *split*, *deletion*, *transposition*, *substitution* dan *insertion*. Kandidat-kandidat kata yang dihasilkan, misalnya "Kam", "Km", "Kmi", "maK", "Kami", "eKm", dan "aKmi". Bahkan kata tersebut akan digabungkan dengan karakter "a" hingga "z". Ketika kandidat kata dihitung peluangnya di korpus, maka kata *typo* "Kmi" adalah kata yang paling mendekati kata "Kami" di korpus. Proses komputasi ini akan terus berjalan pada *background process* hingga ditemukan koreksi kata yang paling mendekati.



Gbr 4. Proses pencarian kandidat kata menggunakan *Peter Norvig Spelling Corrector*

Contoh lainnya seperti yang ditunjukkan pada Tabel I, terdapat kata *typo* “sbelum” (kata yang benar adalah “sebelum”). *Peter Norvig Spelling Corrector* akan mencari kombinasi karakter dalam operasinya untuk mendapatkan kata yang benar dan mencocokkan kata-kata dalam korpus berdasarkan probabilitas.

Pada tahap pemisahan kata (*split*), kata akan dibagi menjadi dua bagian yakni kata kiri dan kanan. Tahap *deletion* akan memanfaatkan hasil *split* dimana kata kiri dipertahankan, tetapi karakter pertama di kata kanan dihapus dan kemudian kata kiri dan kata kanan digabung kembali. Tahap *transposition* juga akan memanfaatkan hasil *split* dimana kata kiri tetap dipertahankan, tetapi posisi karakter pertama di kata kanan ditukar dengan karakter kedua di kata kanan dan kemudian kata kiri dan kata kanan digabung kembali. Pada tahap *substitution*, kata kiri dipertahankan, tetapi posisi karakter pertama pada kata kanan diganti dengan 26 huruf (dari a sampai z) kemudian kata kiri dan kata kanan digabungkan kembali. Pada tahap *insertion*, kata kiri dipertahankan, tetapi posisi karakter pertama pada kata kanan akan disisipkan 26 huruf (dari a sampai z) kemudian kata kiri dan kata kanan digabungkan kembali. Proses pencarian kandidat kata menggunakan *Peter Norvig Spelling Corrector* diilustrasikan pada Gambar 4.

Setelah semua operasi mendapatkan kandidat kata, selanjutnya menghitung peluang masing-masing kandidat kata yang cocok dengan korpus. *Peter Norvig Spelling Corrector* memiliki fungsi korektor kata untuk memilih koreksi ejaan terdekat c untuk kata w . Tidak ada kandidat yang dipilih secara mutlak karena hanya berupa saran perbaikan kata menggunakan probabilitas. Algoritma ini mencoba menemukan koreksi c dari semua kemungkinan koreksi kandidat yang memaksimalkan probabilitas bahwa c adalah koreksi yang ditujukan dengan kata asli w , mengikuti persamaan (1).

$$\operatorname{argmax}_{c \in \text{candidates}} P(c|w)$$

Dengan Teorema Bayes, ini setara dengan persamaan berikut (2):

$$\operatorname{argmax}_{c \in \text{candidates}} \frac{P(c)P(c|w)}{P(w)}$$

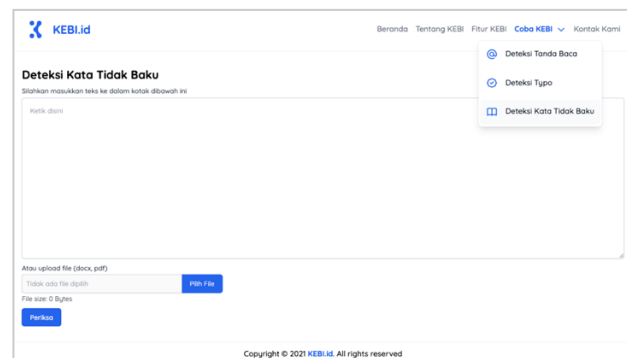
Karena $P(w)$ adalah sama untuk setiap kandidat c yang mungkin, maka dapat memfaktorkannya ke persamaan berikut (3):

$$\operatorname{argmax}_{c \in \text{candidates}} P(c) P(w|c)$$

V. HASIL DAN PEMBAHASAN

Dokumen yang digunakan dalam penelitian ini diperoleh dari karya ilmiah berupa artikel di bidang ilmu komputer yang terbit pada jurnal SANTIKA (Seminar Nasional Informatika) Bela Negara tahun 2020 yang diselenggarakan oleh Program Studi Informatika UPN “Veteran” Jawa Timur. Artikel yang dilibatkan dalam

penelitian ini diperoleh sebanyak 10 judul berbahasa Indonesia. Artikel terdiri dari judul, profil penulis, abstrak, pendahuluan, tujuan, tinjauan pustaka, metode penelitian, pembahasan dan hasil, kesimpulan, ucapan terima kasih, serta referensi. Terdapat 2 skenario dalam penelitian ini, yaitu KEBI 1.0 akan memeriksa kata-kata yang tidak baku dan memeriksa kesalahan pengetikan kata pada 10 artikel tersebut. Evaluasi hasil eksperimen dilakukan dengan menghitung jumlah kata yang terdeteksi dan koreksi ejaan yang akurat. Selain itu, koreksi ejaan yang tidak akurat juga diperhitungkan.



Gbr 5. Antarmuka Pengguna (UI) Aplikasi KEBI 1.0 Checker Berbasis Web

A. Hasil Pengujian Deteksi Kesalahan Ejaan Kata Tidak Baku

Gambar 5 menunjukkan antarmuka aplikasi KEBI 1.0 Checker berbasis Web yang menyediakan beberapa fitur. KEBI dibekali kemampuan untuk mendeteksi kesalahan ejaan berdasarkan tanda baca, kata tidak baku dan kesalahan pengetikan kata (*typo*). Namun, pada penelitian ini terdapat dua fitur utama yang akan dibahas yakni deteksi kata tidak baku dan *typo*. Tabel II menunjukkan bahwa aplikasi memberikan kinerja akurasi terbaik dalam mengoreksi kata tidak baku dengan persentase koreksi kesalahan ejaan 100%.

TABEL II
KINERJA KOREKSI KESALAHAN EJAAN BERDASARKAN KATA TIDAK BAKU MENGGUNAKAN DICTIONARY LOOKUP

Dokumen	Jumlah Halaman	Ukuran File	Koreksi Kesalahan Ejaan Kata Tidak Baku	
			Akurat	Tidak Akurat
Artikel ke-1	5 halaman	173 KB	17 dari 17	0 dari 17
Artikel ke-2	5 halaman	644 KB	12 dari 12	0 dari 12
Artikel ke-3	4 halaman	240 KB	5 dari 5	0 dari 12
Artikel ke-4	6 halaman	142 KB	15 dari 15	0 dari 12
Artikel ke-5	5 halaman	443 KB	6 dari 6	0 dari 12
Artikel ke-6	4 halaman	151 KB	12 dari 12	0 dari 12
Artikel ke-7	6 halaman	167 KB	13 dari 13	0 dari 12
Artikel ke-8	5 halaman	475 KB	15 dari 15	0 dari 12
Artikel ke-9	9 halaman	1.069 MB	72 dari 72	0 dari 12
Artikel ke-10	9 halaman	404 KB	31 dari 31	0 dari 12
Rata-rata (%)			100%	0%

TABEL IV
KINERJA KOREKSI KESALAHAN EJAAN BERDASARKAN KATA TYPO MENGGUNAKAN PETER NORVIG SPELLING CORRECTOR

Dokumen	Jumlah Halaman	Ukuran File	Koreksi Kesalahan Ejaan Kata <i>Typo</i>			
			Akurat	Tidak Akurat	Kata Tidak Bermakna	Istilah Ilmu Komputer
Artikel ke-1	5 halaman	173 KB	0 dari 0	0 dari 0	74	13
Artikel ke-2	5 halaman	644 KB	0 dari 0	0 dari 0	86	54
Artikel ke-3	4 halaman	240 KB	9 dari 15	6 dari 15	38	4
Artikel ke-4	6 halaman	142 KB	6 dari 13	7 dari 13	61	21
Artikel ke-5	5 halaman	443 KB	1 dari 5	4 dari 5	43	53
Artikel ke-6	4 halaman	151 KB	3 dari 5	2 dari 5	49	47
Artikel ke-7	6 halaman	167 KB	20 dari 33	13 dari 33	53	24
Artikel ke-8	5 halaman	475 KB	11 dari 17	6 dari 17	29	1
Artikel ke-9	9 halaman	1.069 MB	14 dari 51	37 dari 51	155	37
Artikel ke-10	9 halaman	404 KB	7 dari 13	6 dari 13	93	73
Rata-rata (%)			49,10%	50,90%	Total = 681	Total = 327

Tabel III menunjukkan daftar kata yang berhasil dikoreksi untuk kesalahan ejaan berdasarkan kata tidak baku yang terdeteksi oleh KEBI 1.0 Checker. Hasil eksperimen menunjukkan bahwa terdeteksi sebanyak 72 kata tidak baku pada artikel ke-9, sedangkan pada artikel ke-3 terdeteksi hanya 5 kata tidak baku. Banyak kesalahan ejaan pada 10 artikel pada eksperimen yang digunakan, seperti kata kerja aktif, keterangan tempat, dan kata baku untuk terminologi umum. Artinya, KEBI 1.0 Checker dapat digunakan untuk mengoreksi kesalahan ejaan pada dokumen eksperimen yang digunakan.

TABEL III
HASIL KOREKSI KESALAHAN EJAAN BERDASARKAN KATA TIDAK BAKU

Kata Tidak Baku	Kata Baku	Kata Tidak Baku	Kata Baku
'syaraf'	'saraf'	'terimakasih'	'terima kasih'
'simple'	'simpler'	'dulu'	'dahulu'
'praktek'	'praktik'	'service'	'servis'
'aktifitas'	'aktivitas'	'diagnosa'	'diagnosis'
'pada hal'	'padahal'	'komplek'	'kompleks'
'disini'	'di sini'	'design'	'desain'
'personil'	'personel'	'familiar'	'familiar'
'analisa'	'analisis'	'test'	'tes'
'mempengaruhi'	'memengaruhi'	'survey'	'survei'
'kedalam'	'ke dalam'	'sub bab'	'subbab'
'resiko'	'risiko'	'standart'	'standar'
'dari pada'	'daripada'	'client'	'klien'
'silahkan'	'silakan'	'stock'	'stok'
'obyek'	'objek'	'didalam'	'di dalam'
...

B. Hasil Pengujian Deteksi Kesalahan Ejaan Kata *Typo*

Pengujian deteksi kata *typo* ini juga menggunakan dokumen dari 10 artikel ilmiah di bidang ilmu komputer. KEBI 1.0 Checker mengoreksi kesalahan ejaan pada 10 artikel ilmiah tersebut menggunakan operasi yang dimilikinya. Tabel IV menunjukkan bahwa aplikasi ini mengoreksi kesalahan ejaan dengan tingkat akurasi 49,10%, sedangkan persentase hasil koreksi kesalahan ejaan yang

tidak akurat sebesar 50,90%. KEBI berhasil mengoreksi kata kerja aktif, kata kerja pasif, kata sifat, kata benda, dan kata baku umum. Penelitian ini menemukan 681 kata yang tidak bermakna dan 327 istilah ilmu komputer yang digunakan pada artikel ilmiah. Kata yang tidak bermakna antara lain seperti nama orang, nama kota, nama jurnal, nama seminar, nama negara, dan singkatan.

TABEL V
HASIL PERBAIKAN KESALAHAN EJAAN BERDASARKAN KATA *TYPO*

Kata <i>Typo</i>	Hasil Perbaikan Kata
'mengembangkan'	'mengembangkan'
'standart'	'standar'
'sikruit'	'sirkuit'
'pemograman'	'pemrograman'
'diletakan'	'diletakkan'
'tebaca'	'terbaca'
'inchi'	'inci'
'penampungan'	'penampungan'
'penelitin'	'penelitian'
'terstuktur'	'terstruktur'
'digunaka'	'digunakan'
'adminitrasi'	'administrasi'
'menghasilkan'	'menghasilkan'
'perputakaan'	'perpustakaan'
'memonitorl'	'memonitor'
'dibubungkan'	'dihubungkan'
'sipanggil'	'dipanggil'
'dibubungkan'	'dihubungkan'
'seringkalai'	'seringkali'
'penelaran'	'penalaran'
'menservice'	'menservis'
'penliti'	'peneliti'
'ataupn'	'ataupun'
'ganggunan'	'gangguan'
'dihudupkan'	'dihidupkan'
'memiliki'	'memiliki'
...	...

Tabel V menunjukkan daftar hasil perbaikan kesalahan ejaan berdasarkan kata *typo* yang telah terdeteksi oleh KEBI 1.0 Checker. Kata-kata di atas merupakan hasil dari

sebagian kata yang dideteksi oleh KEBI 1.0. Tabel VI menunjukkan istilah-istilah komputer yang belum terdeteksi oleh KEBI 1.0 Checker sehingga perlu ditambahkan istilah-istilah seperti nama metode, algoritma, perangkat dan istilah lain yang mewakili bidang ilmu komputer. Sebagai contoh terdapat istilah dalam bentuk singkatan seperti JST (Jaringan Saraf Tiruan), SVM (Support Vector Machine), MQTT (Message Queuing Telemetry Transport), AR (Augmented Reality), VR (Virtual Reality) dan lainnya.

TABEL VI
ISTILAH KOMPUTER YANG BELUM TERDETEKSI OLEH KEBI 1.0 CHECKER

Dokumen	Istilah dalam Ilmu Komputer
Artikel ke-1	JST, SVR, SVM, RSI, EMA, MLR
Artikel ke-2	MQTT, LM, TCP, ESP
Artikel ke-3	ESP, Matlab, Realtime, Uno
Artikel ke-4	EFTS, CDM, PDM, LDM
Artikel ke-5	MRI, VGG, SVM, SVC, InterLR, CoreTM, GHz, Ubuntu, Numpy, RGB, Kaggle, CNN,
Artikel ke-6	Wireless, QoS, Manet, WSN, OLSR, LCD, MPX, DP, mmHg, BPM, GHz, ECG
Artikel ke-7	SMC, CBR, VGA, Hardisk, GHz,
Artikel ke-8	Bit-bit
Artikel ke-9	AR, Anova, PDTO, VR, DOF, SWND, HMD, VLE, Vuforia, Corel, SDK, SPSS
Artikel ke-10	Hadoop, EDA, TWC, Bayes, HDFS, CSV, SQL, XML, JSON, TXT

Tantangan lainnya adalah KEBI belum mampu memperbaiki kata *typo*, misalnya kata yang menunjukkan keterangan tempat. Namun, kata tersebut digabung, misalnya “diatas”, “kedalam”, “disekolah”, “disana”. Kata tersebut seharusnya dipisahkan oleh spasi sehingga yang benar adalah “di atas”, “ke dalam”, “di sekolah”, dan “di sana”.

C. Ukuran Dokumen dan Waktu Pemrosesan

TABEL VII
UKURAN FILE DOKUMEN DAN WAKTU PEMROSESAN KEBI 1.0 CHECKER BERBASIS WEB

Dokumen	Jumlah Halaman	Ukuran File	Waktu Pemrosesan	
			Kata Tidak Baku	Kata <i>Typo</i>
Artikel ke-1	5 halaman	173 KB	0,18 detik	11,36 detik
Artikel ke-2	5 halaman	644 KB	0,32 detik	20,75 detik
Artikel ke-3	4 halaman	240 KB	0,16 detik	17,74 detik
Artikel ke-4	6 halaman	142 KB	0,29 detik	12,55 detik
Artikel ke-5	5 halaman	443 KB	0,24 detik	30,23 detik
Artikel ke-6	4 halaman	151 KB	0,17 detik	10,6 detik
Artikel ke-7	6 halaman	167 KB	0,21 detik	12,56 detik
Artikel ke-8	5 halaman	475 KB	0,18 detik	15,91 detik
Artikel ke-9	9 halaman	1.069 MB	0,3 detik	40,63 detik
Artikel ke-10	9 halaman	404 KB	0,36 detik	44,86 detik
Rata-rata			0,24 detik	21,72 detik

Untuk mengetahui kinerja KEBI 1.0 Checker kepada pengguna, penelitian ini mengevaluasi kinerja aplikasi selama proses pemeriksaan dokumen berdasarkan jumlah

halaman, ukuran file dokumen, dan waktu pemrosesan. Sepuluh artikel yang diujikan pada penelitian terdiri dari 2 hingga 9 halaman, sedangkan ukuran file dari 142 KB hingga 1.069 MB. Tabel VII menunjukkan waktu pemrosesan KEBI 1.0 cukup rendah, rata-rata waktu yang dibutuhkan untuk mengoreksi kata tidak baku adalah 0,24 detik dan untuk mengoreksi kata *typo* adalah 21,72 detik.

VI. KESIMPULAN

KEBI 1.0 Checker menyediakan beberapa fitur yang mampu mendeteksi dan mengoreksi dokumen yang mengandung kata tidak baku dan kesalahan pengetikan kata dalam sebuah karya ilmiah di bidang ilmu komputer. Kata yang tidak baku dideteksi menggunakan *Dictionary Lookup* dan kesalahan pengetikan kata menggunakan *Peter Norvig Spelling Corrector*. Dokumen karya ilmiah yang digunakan dalam eksperimen ini diperoleh dari 10 artikel ilmiah di bidang ilmu komputer yang diterbitkan pada prosiding SANTIKA tahun 2020. Hasil eksperimen menunjukkan bahwa KEBI 1.0 Checker telah memberikan kinerja akurasi terbaik untuk mengoreksi kata-kata tidak baku mencapai 100%, sedangkan untuk mengoreksi kesalahan pengetikan hanya mencapai 49,10%. Penelitian ini menemukan 681 kata yang tidak bermakna dan 327 istilah ilmu komputer yang digunakan pada artikel ilmiah. KEBI berhasil mengoreksi kata kerja aktif, kata kerja pasif, kata sifat, kata benda, dan terminologi umum. Kata yang tidak bermakna seperti nama orang, nama kota, nama jurnal, nama seminar, nama negara, dan singkatan. KEBI belum mampu mengoreksi istilah-istilah yang digunakan dalam ilmu komputer seperti nama metode, algoritma, perangkat dan istilah lainnya. Untuk menguji aplikasi, penelitian ini menggunakan 10 artikel ilmiah yang terdiri dari 2 hingga 9 halaman, sedangkan ukuran file mulai dari 142 KB hingga 1.069 MB. Waktu pemrosesan KEBI 1.0 cukup rendah, rata-rata waktu yang dibutuhkan untuk mengoreksi kata tidak baku adalah 0,24 detik dan untuk mengoreksi kata yang salah ketik adalah 21,72 detik. Untuk penelitian selanjutnya, KEBI 1.0 Checker membutuhkan korpus yang besar untuk menangani kata *typo* dari berbagai bidang karya ilmiah seperti ilmu komputer, teknik, ekonomi, sosial, kedokteran dan kosakata lainnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kementerian Pendidikan dan Kebudayaan Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah mendanai penelitian ini berdasarkan Surat Perjanjian Penugasan Dalam Rangka Pelaksanaan Program Penelitian Internal Batch I untuk Skema Riset Dasar (Risda) Universitas Pembangunan Nasional “Veteran” Jawa Timur Tahun Anggaran 2021, Nomor: SPP / 12 / UN.63.8/LT/IV/2021.

REFERENSI

- [1] S. Ariana, "Kesalahan Penggunaan Ejaan yang Disempurnakan dalam Karya Ilmiah Dosen Universitas Bina Darma," *Jurnal Ilmiah Bina Edukasi*, vol. 5, no. 5, pp. 53-62, 2011.
- [2] K. M. Hindayani, I. G. S. M. Diyasa, P. A. Riyantoko dan T. M. Fauruddin, "Studi Literatur Mengenai Prediksi Harga Saham Menggunakan Machine Learning," dalam *Seminar Nasional Informatika Bela Negara (SANTIKA)*, Surabaya, 2020.
- [3] H. Alwi dan dkk, Tata Bahasa Baku Bahasa Indonesia, Jakarta: Balai Pustaka, 1998.
- [4] Murtiningsih, "Kesalahan Berbahasa Indonesia Mahasiswa S-1 PGSD STIKIP Nuwar Fak-fak," *Peneliti Ilmu Pendidik*, vol. 6, no. 1, pp. 74-82, 2013.
- [5] G. L. Y. Londo, Y. S. P. W.P. dan M. Maslim, "Pembangunan Aplikasi Identifikasi Kesalahan Ketik Dokumen Berbahasa Indonesia Menggunakan Algoritma Jaro-Winkler Distance," *AKSIS: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, vol. 2, no. 2, pp. 138-153, 2018.
- [6] T. Hartina dan A. Masri, "Pendeteksi Kesalahan Pengetikan Kata Nonbaku pada Karya Tulis Menggunakan N-Gram," *Jurnal Informatika*, vol. 7, no. 1, pp. 77-84, 2020.
- [7] R. N. E. Anggraini, M. A. Zinni dan S. Rochimah, "Kakas Bantu Pendeteksi Kesalahan Tanda Baca pada Karya Tulis Ilmiah," *JUTI*, vol. 14, no. 1, pp. 117-125, 2016.
- [8] D. Surianto, D. Triyanto dan U. Ristian, "Penerapan Algoritma Boyer Moore dan Metode N-Gram pada Aplikasi Penyunting Naskah Teks Bahasa Indonesia Berbasis Web," *Coding: Jurnal Komputer dan Aplikasi*, vol. 8, no. 3, pp. 50-60, 2020.
- [9] Y. P. Putri dan R. Lawson, "Aplikasi Pengkoreksi Kesalahan Ejaan dan Padanan Kata pada Tugas Akhir Mahasiswa," *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, vol. 14, no. 2, pp. 72-75, 2019.
- [10] I. R. R. H. Junar Frando, "Penerapan Jaro Winkler Distance dalam Aplikasi Pengkoreksi Kesalahan Penulisan Bahasa Indonesia Berbasis Web," *Coding, Jurnal Komputer dan Aplikasi*, vol. 7, no. 3, pp. 44-53, 2019.
- [11] I. C. A. W. W. Tusty Nadia Maghfira, "Deteksi Kesalahan Ejaan dan Penentuan Rekomendasi Koreksi Kata yang Tepat Pada Dokumen Jurnal JTIK Menggunakan Dictionary Lookup dan Damerau-Levenshtein Distance," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 498-506, 2017.
- [12] A. Azahari, Teknik Penulisan Ilmiah. 3rd ed., Tangerang Selatan: Universitas Terbuka, 2019.
- [13] T. P. P. B. Indonesia, Pedoman Umum Ejaan Bahasa Indonesia, Jakarta: Badan Pengembangan dan Pembinaan Bahasa Kementerian Pendidikan dan Kebudayaan, 2016.
- [14] V. S. Ningrum, "Penggunaan Kata Baku dan Tidak Baku di Kalangan Mahasiswa," *Jurnal Pembelajaran Bahasa dan Sastra Indonesia*, vol. 5, no. 2, pp. 22-27, 2019.
- [15] A. N. Astawijaya dan K. K. Purnamasari, "Perbandingan Levenshtein, Smith-Waterman, Needleman-Wunsch dalam Typo Checking," *Jurnal Ilmiah Komputer dan Informatika (Komputa)*, 2019.
- [16] S. Mujilawati, "Pre-processing Text Mining pada Data Twitter," dalam *Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016)*, Yogyakarta, 2016.
- [17] S. dan S. Saudah, Buku Ajar Bahasa Indonesia Akademik, Yogyakarta: Pustaka Pelajar, 2015.