

ABSTRAK

Penggunaan teks informal pada media digital, khususnya media sosial, sering menimbulkan kesalahan ejaan seperti typo, penggunaan kata tidak baku, dan penulisan kata tanpa spasi yang dapat menurunkan kualitas informasi serta menyulitkan proses pemahaman teks. Oleh karena itu, diperlukan sistem spelling correction yang mampu memperbaiki kesalahan tersebut secara otomatis pada teks Bahasa Indonesia. Penelitian ini mengimplementasikan metode *Damerau Levenshtein Distance* (DLD) dan metode Empiris untuk memperbaiki kesalahan ejaan pada teks informal. Metode DLD digunakan untuk mengukur tingkat kemiripan antara kata salah dan kata baku berdasarkan kamus KBBI melalui perhitungan edit distance, sedangkan metode Empiris digunakan untuk menangani kesalahan penulisan berupa penggabungan kata tanpa spasi. Tahapan penelitian meliputi preprocessing data berupa *cleaning data*, *case folding*, dan *tokenizing*, kemudian dilanjutkan dengan proses pengecekan kata menggunakan kamus serta proses koreksi kata menggunakan pendekatan Top-3 kandidat. Dataset yang digunakan terdiri dari 2000 ulasan layanan kesehatan berbahasa Indonesia. Pengujian dilakukan pada tiga skenario, yaitu penggunaan metode DLD, metode Empiris, dan kombinasi metode DLD dengan Empiris. Hasil pengujian menunjukkan bahwa metode DLD memperoleh akurasi sebesar 37,7%, metode Empiris memperoleh akurasi sebesar 51,37%, sedangkan kombinasi metode DLD dan Empiris menghasilkan akurasi terbaik sebesar 84,4%. Hasil tersebut menunjukkan bahwa metode Empiris mampu membantu proses koreksi pada kasus penggabungan kata, sedangkan kombinasi kedua metode mampu memberikan performa yang lebih optimal dalam menangani kesalahan karakter maupun kesalahan struktur kata pada teks informal Bahasa Indonesia.

Kata Kunci : *Damerau Levenshtein Distance*, Empiris, *Spelling correction*

ABSTRACT

The use of informal text in digital media, especially social media, often causes spelling errors such as typos, non-standard words, and words written without spaces, which can reduce the quality of information and make text comprehension more difficult. Therefore, a spelling correction system is needed to automatically correct such errors in Indonesian text. This study implements the Damerau Levenshtein Distance (DLD) method and the Empirical method to correct spelling errors in informal text. The DLD method is used to measure the similarity between incorrect words and standard words based on the Indonesian Dictionary (KBBI) through edit distance calculation, while the Empirical method is used to handle errors involving concatenated words without spaces. The research stages include data preprocessing consisting of data cleaning, case folding, and tokenizing, followed by word validation using a dictionary and word correction using the Top-3 candidate approach. The dataset used consists of 2,000 Indonesian healthcare service reviews. Testing was conducted in three scenarios: using the DLD method, the Empirical method, and the combination of DLD and Empirical methods. The test results show that the DLD method achieved an accuracy of 37.7%, the Empirical method achieved an accuracy of 51.37%, while the combination of DLD and Empirical methods produced the best accuracy of 84.4%. These results indicate that the Empirical method is able to assist the correction process in cases of concatenated words, while the combination of both methods provides more optimal performance in handling character errors and word structure errors in informal Indonesian text.

Keywords: *Damerau Levenshtein Distance, Empirical, Spelling Correction*