

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa Indonesia merupakan bahasa yang digunakan oleh rakyat Indonesia dalam berkomunikasi. Bahasa Indonesia menjadi identitas bangsa di tengah-tengah bangsa lain di dunia. Sebelum resmi menjadi bahasa nasional, bahasa Indonesia dikenal sebagai bahasa Melayu, pada tanggal 25–28 Juni 1938, Kongres bahasa Indonesia I di Solo. Dari hasil Kongres bahasa Indonesia I ini dapat disimpulkan bahwa usaha pembinaan dan pengembangan bahasa Indonesia telah dilakukan secara sadar oleh cendekiawan dan budayawan kita. Sejak tanggal 18 Agustus 1945, bahasa Indonesia dipakai resmi oleh bangsa Indonesia sebagai bahasa nasional [1]. Sejak tahun 1945 bahasa Indonesia mengalami perkembangan yang sangat pesat sebagai dampak kemajuan ilmu pengetahuan, teknologi dan seni penggunaannya pun semakin luas dalam beragam ranah pemakaian baik secara lisan maupun tulis.

Di era saat ini perkembangan teknologi semakin canggih, masyarakat mulai banyak menggunakan komputer, *smartphone* dan *internet*, budaya tulis menulis juga terus berkembang. Hanya saja penggunaan kertas mulai berkurang dan beralih pada tulis menulis menggunakan media-media canggih seperti komputer dan *smartphone* [2]. Seringkali pengguna melakukan kesalahan dalam penulisan ejaan kata bahasa Indonesia. Bahasa yang digunakan harus mengikuti kaidah-kaidah kebahasaan yang ada. Kesalahan penggunaan bahasa bisa menimbulkan interpretasi yang berbeda antara orang yang satu dan yang lainnya. Bertolak dari hal tersebut, penulis merasa penting untuk menganalisis kesalahan penggunaan ejaan bahasa Indonesia (EBI) [3]. Kesalahan penulisan atau ejaan yang sering terjadi disebabkan oleh dua hal yaitu kelalaian pengguna dan kata tersebut belum ada di dalam kamus [4]. Kesalahan-kesalahan yang umumnya sering terjadi antara lain : penggantian satu huruf, penyisipan satu huruf, penghilangan satu huruf, maupun penukaran dua huruf berdekatan.

Dalam membantu proses pengoreksian kesalahan ejaan kata diperlukan sebuah perangkat lunak yang dapat membantu pekerjaan untuk mengedit naskah tulisan. Apabila ditemukan kata yang salah maka harus dilakukan pencarian kemungkinan

kata yang sesuai. Pada proses pencarian kemungkinan kata tersebut, diperlukan suatu pendekatan pencarian *string* khusus yaitu dengan pencarian jarak terkecil [4].

Dalam pencarian *string* dibutuhkan sebuah metode yang dapat mencari jarak terkecil, dalam hal ini Algoritma *Levenshtein Distance* mampu mencari jarak dari kedua *string* [5]. *Levenshtein distance* adalah sebuah matriks *string* yang digunakan untuk mengukur perbedaan atau jarak (*distance*) antara dua *string*. Nilai *distance* antara dua *string* ini ditentukan oleh jumlah minimum dari operasi-operasi perubahan yang diperlukan untuk melakukan transformasi dari suatu *string* menjadi *string* lainnya. Operasi-operasi tersebut adalah penyisipan (*insertion*), penghapusan (*deletion*), atau penukaran (*substitution*). *Levenshtein distance* merupakan salah satu algoritma yang dapat digunakan dalam mendeteksi kemiripan antara dua *string* [6].

Dalam penelitian sebelumnya yang dilakukan oleh Rachmania, dkk [7], pengoreksian ejaan kata berbahasa Indonesia dengan menggunakan algoritma *Levenshtein Distance*. Penelitian yang dilakukan oleh Dewi, dkk [8], melakukan penelitian pendeteksian ejaan kata bahasa Indonesia yang dilakukan dengan proses pendekatan *string* dan pencocokan kata pada *database*. Dalam penelitiannya, menggunakan metode *Approximate String Matching* dengan algoritma *Levenshtein Distance* [8], dan dalam penelitian [4], untuk mendeteksi kesalahan penulisan ejaan bahasa Indonesia menggunakan algoritma *Levenshtein Distance* [4]. Dari penelitian sebelumnya hasil pengujian dengan menggunakan algoritma *Levenshtein Distance* hasil pencarian sugesti bahwa sistem sudah bisa mendeteksi kesalahan dan memberi sugesti untuk tiap kategori kesalahan ejaan menghasilkan keakuratan sebesar yang cukup baik [4][7][8] dari beberapa penelitian sebelumnya terdapat kekurangan dari segi tingkat akurasi dan waktu pemrosesan. Namun untuk algoritma *Levenshtein Distance* telah banyak dikembangkan seiring berjalannya waktu. Dengan banyak perkembangan yang telah terjadi dipilihnya pembaharuan dari Algoritma *Levenshtein Distance* yang dimana merupakan algoritma *Damerau Levenshtein Distance*.

Diharapkan dari penelitian ini algoritma *Damerau Levenshtein Distance* menghasilkan tingkat akurasi yang baik pada pencarian sugesti kata pada ejaan yang kurang tepat.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Kesalahan umum yang sering terjadi oleh pengguna saat menginputkan kata ialah kesalahan ejaan dalam penulisan. Hal ini dapat menimbulkan interpretasi yang berbeda antara orang yang satu dengan yang lainnya, dalam hal ini diperlukan sebuah metode dimana dapat mengatasi beberapa masalah yang sering terjadi oleh pengguna di saat menginputkan kata.

1.2.2 Metode Usulan

Diusulkan Algoritma *Damerau Levenshtein Distance* untuk memperbaiki ejaan kata bahasa Indonesia.

1.2.3 Pertanyaan Penelitian

Seberapa akurat *Algoritma Damerau Levenshtein Distance* untuk memperbaiki ejaan kata bahasa Indonesia yang kurang tepat.?

1.3 Tujuan Dan Manfaat

1.3.1 Tujuan Penelitian

Diharapkan Dengan permasalahan yang sering terjadi diharapkan metode yang digunakan mampu memperbaiki kesalahan ejaan dalam menginputkan kata.

1.3.2 Manfaat Penelitian

Penelitian ini diharapkan bermanfaat bagi pengguna sistem dan pengembang sistem. Bagi pengguna sistem bermanfaat untuk koreksi ejaan bahasa Indonesia dan bagi pengembang sistem dapat digunakan sebagai fitur tambahan dalam sistem yang lebih besar.

1.4 Batasan Masalah

1. Aplikasi ini hanya mampu mengoreksi kata dasar yang terdapat dalam kamus besar bahasa Indonesia (KBBI) [16], dan korpus kata imbuhan bahasa Indonesia [17].

2. Sistem ini akan dibangun menggunakan bahasa pemograman PHP.
3. Perbaikan kata dilakukan hanya pada struktur kata dasar yang kurang tepat dimana sistem tidak dapat mengoreksi struktur kata imbuhan yang kurang tepat.