

BAB I

PENDAHULUAN

1.1 LatarBelakang

Bahasa adalah salah satu komponen yang paling penting dalam kehidupan manusia. Dalam bentuk tulisan, bahasa menyimpan pengetahuan dari satu generasi ke generasi lain. Sedangkan dalam bentuk lisan, bahasa berperan dalam mengarahkan tingkah laku manusia sehari-hari dalam berhubungan dengan orang lain. Bahasa menjadi faktor penting pula dalam penulisan dokumen. Apabila dalam penulisan dokumen terdapat tulisan yang salah, maka artinya akan menjadi berbeda. Pada era sekarang ini penulisan dokumen banyak menggunakan komputer, dan pada komputer tidak selalu ada program yang dapat mendeteksi kesalahan pada penulisan bahasa. Kesalahan pengetikan dokumen memang sering sekali terjadi, apalagi belakangan ini kesadaran masyarakat untuk menuangkan idenya ke dalam artikel, jurnal ilmiah, tugas kuliah ataupun dokumen lainnya mengalami peningkatan. Tentu saja dalam penulisan dokumen tersebut adanya kesalahan pengetikan yang disebabkan oleh beberapa faktor seperti, letak huruf pada *keyboard* yang berdekatan, dan kesalahan karena kegagalan mekanis atau slip dari tangan atau jari. Oleh karena itu, dibutuhkan sebuah program yang dapat mendeteksi kesalahan penulisan bahasa.

Fasilitas pengolah kata (pemeriksa ejaan / *spell checker*) adalah sebuah fasilitas yang memungkinkan pengguna aplikasi pengolah kata memeriksa ejaan pada semua dokumen yang diketik serta memberikan usulan kata-kata untuk yang salah ejaannya. Hal ini dapat meminimumkan kemungkinan salah eja atau salah ketik. Di negara-negara maju penggunaan fasilitas pengolah kata elektronik sangat umum, sehingga menjadi salah satu indikator pemilihan terhadap pengolah kata yang hendak dipakai. Pengetikan dengan cara manual akan menghabiskan banyak waktu dan membutuhkan suatu sumber pasti sebagai acuan bahwa kata tersebut memang salah dalam proses penulisannya. Efisiensi waktu yang dibutuhkan jika dilakukan dengan manual tentunya tidak akan optimal dan cukup

membosankan sehingga kemungkinan adanya kesalahan manusia dapat mengakibatkan proses pengecekan kata menjadi tidak optimal. Dari beberapa algoritma dapat diimplementasikan dalam memberikan kata saran yang paling mendekati dari kata yang salah penyetikannya. Salah satunya adalah Algoritma *Levenshtein Distance* yang dapat menghitung jarak keterbedaan antara 2 (dua) String (Andhika,2010) .berdasarkan hasil pengujian pencarian kata, sistem bisa mendeteksi 100% kesalahan kata (Rachmania Nur Dwitiyastuti).

Levenshtein Distance, atau sering disebut juga sebagai *edit distance*, adalah suatu pengukuran (metrik) yang dihasilkan melalui perhitungan jumlah perbedaan yang terdapat pada dua string. *Levenshtein distance* antara dua string didefinisikan sebagai jumlah minimum perubahan yang diperlukan untuk mengganti suatu string dengan string lain, dengan operasi penambahan (*insert*), penghapusan (*delete*), atau penggantian karakter (*substitute*) pada suatu karakter. Yang dimaksud dengan *distance* adalah jumlah perubahan yang diperlukan untuk mengubah suatu bentuk string ke bentuk string yang lain. Contohnya, string “hasil” dan “hasal” memiliki *distance* 1 karena diperlukan satu operasi untuk mengubah string “hasal” menjadi “hasil”.

Hasil *Levenshtein distance* yang diperoleh sebenarnya tidak dapat langsung dimanfaatkan, namun perlu diolah untuk memenuhi kebutuhan aplikasi tersebut. Banyak aplikasi yang menggunakan algoritma ini, seperti pengecek ejaan, pemandu penerjemahan, perkiraan dari pengucapan dialek, mesin pencari, pemberi revisi file dengan membandingkan perbedaan dua buah file, pendeteksi pemalsuan, pengenalan percakapan (*speech recognition*), dan sebagainya.

Dalam tugas akhir ini penulis akan mengimplementasikan algoritma *Levenshtein Distance* yang dapat membantu memeriksa kesalahan penulisan pada sebuah dokumen berbahasa Indonesia.

1.2 Rumusan Masalah

Penelitian ini merumuskan beberapa masalah, yakni:

1. Bagaimana membuat sebuah program *Spell Checker* (pemeriksa ejaan) dengan menerapkan algoritma *Levenshtein Distance* untuk kata dari bahasa indonesia ?
2. Bagaimana kinerja algoritma *Levenshtein Distance* untuk pemeriksaan ejaan pada sebuah dokumen berbahasa indonesia ?

1.3 Batasan Masalah

Beberapa batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Bahasa yang ditetapkan sebagai dasar evaluasi penulisan dalam program yang dibangun adalah bahasa indonesia.
2. Format dokumen yang di gunakan adalah format *.txt*.

1.4 Tujuan Penelitian

1. Mengimplementasikan algoritma *Levenshtein Distance* pada sebuah program *spelling checker* (pemeriksa ejaan) .
2. Mengukur kinerja algoritma *Levenshtein Distance* untuk *spelling checker* (pemeriksa ejaan) pada dokumen berbahasa indonesia .

1.5 Manfaat Penelitian

Membantu dalam mengurangi salah pengetikan pada sebuah dokumen berbahasa Indonesia yang dikarenakan salah pengetikan agar sesuai dengan kaidah penulisan Bahasa Indonesia dengan ketelitian yang cukup tinggi, serta dapat membantu dalam pembuatan naskah karya ilmiah maupun skripsi.