

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Bahasa adalah alat komunikasi lingual baik secara lisan maupun tulisan. Kesalahan pengetikan adalah kesalahan penekanan tombol keyboard sehingga kata yang diketik tidak sama dengan apa yang seharusnya diketik. Kesalahan dalam pengetikan umumnya disebabkan karena kedekatan letak keyboard, tombol keyboard rusak, slip jari, pengetikan yang tergesa-gesa, dan lain-lain. Kesalahan dalam pengetikan dapat membuat penerima pesan tidak mendapatkan maksud dari apa yang disampaikan.

Beberapa penelitian tentang pengkoreksian kesalahan pengetikan kata bahasa Indonesia sudah dilakukan sebelumnya. Penelitian Wedha Wardhana dan kawan-kawan menggunakan metode *n-gram* dan memberikan kesimpulan *n-gram* kurang baik mengoreksi kesalahan yang disebabkan oleh substitusi (huruf yang salah) dan transposisi (huruf yang tertukar) [1]. Penelitian Norfa Yoga menggunakan Metode *Double Metaphone* menghasilkan akurasi yang kurang baik. Penelitian tersebut memberi kesimpulan bahwa hasil pendeteksian kesalahan kata bergantung pada kelengkapan kata pada kamus data yang digunakan [2].

Salah satu algoritma yang populer untuk pengkoreksian kesalahan pengetikan adalah Algoritma *Levenshtein*. Salah satu penelitian yang berkaitan dengan Algoritma *Levenshtein* antara lain penelitian Muhammad Maulana dan kawan-kawan [3] yang menggunakan Algoritma *Levenshtein Distance* untuk pengecekan ejaan bahasa Inggris dan memberikan hasil yang baik. Penelitian Febi [4] melakukan optimasi *Levenshtein* dengan *Dynamic Programming* karena solusi pencarian *Levenshtein Distance* mengakibatkan waktu eksekusi menjadi sangat lama ketika kamus yang digunakan sangat banyak. Penelitian Peter Norvig [5] menggunakan pendekatan yang berbeda dengan pendekatan tradisional seperti *Levenshtein Distance* metode tersebut menghitung poin dari setiap kata di kamus dan mengurutkannya. Penelitian Peter Norvig menghasilkan semua kemungkinan kata dengan melakukan penghapusan, penukaran, penyisipan, dan penggantian

huruf dan mengeceknya dikamus, sehingga pendekatan ini bisa lebih cepat. Penelitian lainnya adalah Wolf Garbe [6] yang menciptakan Algoritma *SymSpell* (*Symmetric Delete Spelling Correction*) yang berbasis dari penelitian Peter Norvig dan berhasil meningkatkan performanya menjadi seribu kali lebih cepat. Algoritma *SymSpell* mengurangi kompleksitas dari pencarian kamus dengan melakukan penghapusan saja. Berbeda dengan pendekatan Peter Norvig yang melakukan penghapusan, penukaran, pengisipan, dan penggantian huruf di kata yang akan dicek, pendekatan Wolf Garbe melakukan penghapusan saja namun tidak hanya pada kata yang akan dicek namun kata yang ada di kamus, sehingga bisa lebih cepat.

Kasus yang akan diangkat dalam penelitian ini adalah kesalahan ejaan kata pada *question answering* berbahasa Indonesia. *Question Answering* bertujuan untuk menampilkan jawaban atas pertanyaan yang diberikan pengguna. *Question Answering System* dapat diterapkan pada *front office* untuk menangani pertanyaan yang diajukan pengunjung. Contohnya pada front office UNIKOM, pengunjung sering bertanya mengenai informasi pendaftaran atau penerimaan mahasiswa baru. Pengunjung dapat menanyakan hal-hal sederhana seperti besarnya biaya yang harus dibayarkan, kapan dimulainya ujian saringan masuk, yang seharusnya dapat dijawab dengan cepat. Pendeteksian dan pengkoreksian kesalahan kata dapat diterapkan di *question answering system* agar ketika pengguna melakukan kesalahan dalam pengetikan kata, sistem bisa memberikan rekomendasi kata dan diharapkan jawabannya tetap relevan.

Hasil pendeteksian dan pengkoreksian kesalahan kata bergantung pada kelengkapan kata pada kamus kata yang digunakan, semakin lengkap kamus yang digunakan maka akan semakin banyak kata yang ada. Penggunaan algoritma *SymSpell* diharapkan dapat memberikan akurasi dan juga performa yang baik dalam perbaikan kesalahan pengetikan kata bahasa Indonesia. Berdasarkan permasalahan tersebut, penelitian ini akan membahas tentang “PERBAIKAN KESALAHAN EJAAN DENGAN METODE SYMSPELL PADA KASUS TANYA JAWAB DALAM BAHASA INDONESIA”.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, maka rumusan masalah dari penelitian ini adalah bagaimana cara memperbaiki kesalahan ejaan dengan Algoritma *SymSpell* dalam pengetikan kata bahasa Indonesia di *Question Answering System*.

1.3. Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk memperbaiki kesalahan ejaan dengan Algoritma *SymSpell* dalam pengetikan kata bahasa Indonesia di *Question Answering System*. Sedangkan tujuan dari penelitian ini adalah untuk mengukur akurasi dan performa Algoritma *SymSpell* untuk pendeteksi dan pengkoreksi kesalahan dalam pengetikan kata bahasa Indonesia di *Question Answering System*.

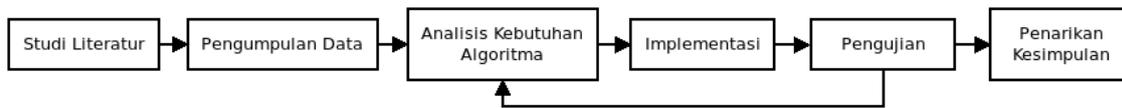
1.4. Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang diproses adalah kata dalam bahasa Indonesia.
2. Jenis *typo* yang diproses adalah *non-word* yaitu disebabkan oleh ejaan yang tidak terdapat pada kamus.
3. Algoritma *SymSpell* diterapkan pada *question answering system*.
4. Sistem akan memunculkan rekomendasi kata saat kata yang diketik pengguna terdeteksi salah.
5. Kasus di *question answering system* adalah *front office* UNIKOM tentang pembayaran kuliah.

1.5. Metode Penelitian

Metode penelitian adalah tahapan atau cara teratur untuk mencapai tujuan dari penelitian yang dilakukan. Metode penelitian pada penelitian ini terdiri dari beberapa tahap seperti terlihat pada Gambar 1.1 berikut.



Gambar 1. 1Metode Penelitian

1) Studi Literatur

Studi literatur dilakukan untuk mendapatkan pengetahuan tentang Algoritma *SymSpell* yaitu dengan mengumpulkan informasi dari buku, jurnal, paper, dan pencarian materi di *internet*.

2) Pengumpulan Data

Mengumpulkan data yang dibutuhkan untuk proses analisis dan pembangunan perangkat lunak. Contohnya data kamus dan data korpus.

3) Analisis Kebutuhan Algoritma

Menganalisis kebutuhan algoritma dari Algoritma *SymSpell* dan *Rule-Based* untuk *Question Answering System* mulai dari masukan, cara kerja, sampai keluaran yang dihasilkan. Juga dilakukan analisis terhadap kebutuhan fungsional dan perancangan dengan UML serta kebutuhan non-fungsional.

4) Pembangunan Perangkat Lunak

Mengimplementasikan hasil analisis ke dalam program agar selanjutnya bisa masuk ke tahap pengujian.

5) Pengujian

Setelah tahap implementasi selanjutnya dilakukan pengujian untuk mengetahui hasil dari tahap implementasi. Jika pengujian memberikan hasil yang kurang baik maka akan kembali ke tahap Analisis Kebutuhan Algoritma untuk menganalisis kesalahan dari cara kerja atau parameter yang ditentukan.

6) Penarikan Kesimpulan

Setelah tahap pengujian selesai maka bisa ditarik kesimpulan berdasarkan tujuan penelitian yang telah ditentukan sebelumnya.

1.6. Sistematika Penulisan

Sistematika penulisan ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan secara umum adalah sebagai berikut:

BAB 1 PENDAHULUAN

Berisi penjelasan tentang latar belakang mengapa penelitian ini dilakukan, rumusan masalah berdasarkan latar belakang, maksud dan tujuan, batasan-batasan dari penelitian yang dilakukan, serta metodologi penelitian yang digunakan.

BAB 2 LANDASAN TEORI

Berisi teori dan konsep yang digunakan sebagai landasan atau acuan yang diperlukan yaitu antara lain, penjelasan tentang Algoritma *SymSpell*, serta penjelasan singkat mengenai *Unified Modeling Language* dan *Java Programming Language*.

BAB 3 ANALISIS KEBUTUHAN ALGORITMA

Berisi pembahasan analisis dari rumusan masalah yang sudah dijelaskan di bagian pendahuluan, analisis kebutuhan algoritma mulai dari analisis masalah, analisis *input*, analisis metode, sampai analisis kebutuhan fungsional dan *non-fungsional*.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Berisi pembahasan implementasi yang dilakukan, mulai dari bahasa pemrograman, *hardware* dan *software* yang digunakan, serta hasil pengujian. Pengujian yang dilakukan antara lain pengujian akurasi dan pengujian performa.

BAB 5 KESIMPULAN DAN SARAN

Berisi kesimpulan dari hasil pengujian yang dilakukan, apakah tujuan awal dalam penelitian telah tercapai atau tidak, serta berisi saran untuk penelitian selanjutnya.

