

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Pada era sekarang ini penulisan dokumen banyak yang menggunakan komputer, dan pada setiap komputer tidak selalu ada program yang dapat mendeteksi kesalahan penulisan [1]. Kesalahan dalam penulisan dapat disebabkan oleh kesalahan dalam pengetikan dokumen yang tidak sesuai dengan ejaan kata yang tepat yang dapat mengubah kata menjadi tidak bermakna (non-word) dan dapat mengubah menjadi kata yang memiliki makna berbeda (real-word) [2]. Apabila dalam pengetikan dokumen terdapat tulisan yang salah, maka dapat mengubah arti dan mengubah pemahaman pembacanya.

Pada penelitian yang berkaitan dengan *typo checking* sebagai solusi dari kesalahan penulisan kata terdapat penelitian *typo checking* pada bahasa India. Kesimpulan yang didapat pada penelitian tersebut setelah mengambil 2 *sample* yaitu, algoritma *Smith Waterman* mendapat nilai akurasi terbaik, lalu algoritma *Needleman Wunsch* mendapat nilai akurasi pertengahan dan algoritma *Levenshtein* kurang baik untuk memperbaiki kesalahan dalam pengetikan [3]. Sedangkan dalam penelitian *typo checking* dalam bahasa Indonesia, kesimpulan yang didapat adalah penggunaan *Levenshtein Distance* untuk memperbaiki kesalahan dalam pengetikan sudah tepat dan dapat memberikan hasil akurasi yang cukup tinggi yaitu 86,4% [4].

Pada penelitian ini akan menggunakan algoritma *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch*, karena dibutuhkan algoritma mana yang mendapatkan solusi yang optimal selain *Levenshtein* sebagai pembandingan. Berdasarkan uraian di atas, maka pada penelitian ini akan dilakukan *typo checking* dengan membandingkan algoritma *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch*, untuk menentukan akurasi pada kasus pendeteksian kesalahan penulisan kata pada teks berita Berbahasa Indonesia.

1.2. Identifikasi Masalah

Berdasarkan uraian latar belakang masalah yang telah dijelaskan, maka identifikasi masalah yang ada adalah dibutuhkan hasil perbandingan akurasi pendeteksian kesalahan kata yang menggunakan metode *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch* pada Bahasa Indonesia agar mendapat solusi terbaik untuk *typo checking*.

1.3. Maksud dan Tujuan

Penelitian ini bermaksud membangun sistem pendeteksian kesalahan kata pada teks berbahasa Indonesia dengan tujuan yang ingin dicapai dari penelitian ini adalah mengetahui akurasi pendeteksian kesalahan kata dengan membandingkan metode *Levenshtein*, *Smith-Waterman* dan *Needleman-Wunsch*.

1.4. Batasan Masalah

Batasan masalah adalah pembatasan ruang lingkup pekerjaan yang akan dilakukan terhadap permasalahan yang ada agar pembahasan menjadi lebih jelas dan terarah. Batasan dalam penelitian ini adalah sebagai berikut.

Masukan :

1. Dokumen yang diproses adalah teks Berbahasa Indonesia.
2. Jenis dokumen teks adalah dokumen tunggal yang berformat .txt.
3. Data uji berasal dari artikel dan berita yang bersumber dari Internet seperti kompas.com [6].
4. Koreksi kesalahan kata dibatasi hanya pada kesalahan ejaan non-word.

Proses :

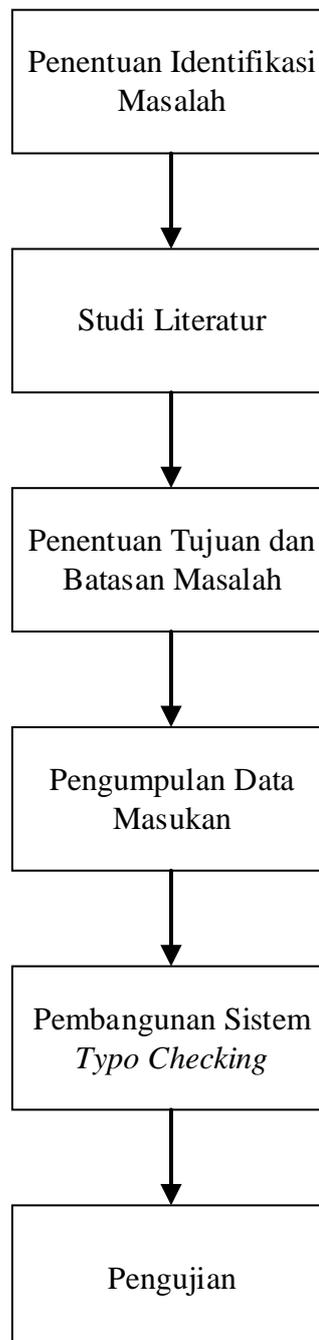
1. Preprocessing yang digunakan yaitu: case folding, filtering dan tokenisasi kata.
2. Proses menentukan kata yang dianggap salah ketik.
3. Proses memberikan rekomendasi kata yang dianggap salah ketik dan perbaikan kata.

Keluaran :

Keluaran yang dihasilkan berupa hasil deteksi kata yang dianggap salah ketik, dan kata yang terpilih sebagai perbaikan kata yang dianggap salah ketik.

1.5. Metode Penelitian

Penelitian ini menggunakan metode deskriptif dikarenakan penelitian ini mengacu dari penelitian yang telah dilakukan sebelumnya untuk mendapatkan gambaran yang jelas mengenai hal – hal apa saja yang diperlukan untuk penelitian ini [7].



Gambar 1.1 Skema Metode Penelitian [7]

1. Penentuan Identifikasi Masalah

Menentukan identifikasi masalah pada penelitian merupakan tahapan dimana peneliti mengidentifikasi dan merumuskan apa – apa saja yang menjadi masalah.

2. Studi Literatur

- a. Pencarian dan pengumpulan kajian-kajian yang berkaitan dengan masalah-masalah yang ada pada penelitian ini, baik berupa artikel, buku referensi, mencari referensi via internet dan sumber-sumber lain yang berhubungan dengan masalah tugas akhir. Termasuk melakukan bimbingan dengan dosen pembimbing.
- b. Pengumpulan data-data pendukung dan melakukan Analisa (spesifikasi) sistem yang diperlukan dalam tahap implementasi.

3. Penentuan Tujuan dan Batasan Masalah

Menentukan tujuan dan batasan masalah penelitian merupakan tahapan dimana peneliti merumuskan tujuan apa yang ingin dicapai dan apa saja batasan masalah yang menjadi acuan dalam penelitian ini seperti dokumen yang diproses adalah teks. Berbahasa Indonesia dan koreksi kesalahan kata dibatasi hanya pada kesalahan ejaan non-word

4. Pengumpulan Sumber Data Masukan

Mencari dan mengumpulkan data untuk inputan merupakan tahapan untuk mengumpulkan data – data yang diperlukan pada penelitian ini. Dalam penelitian ini..

5. Pembangunan Sistem *Typo Checking*

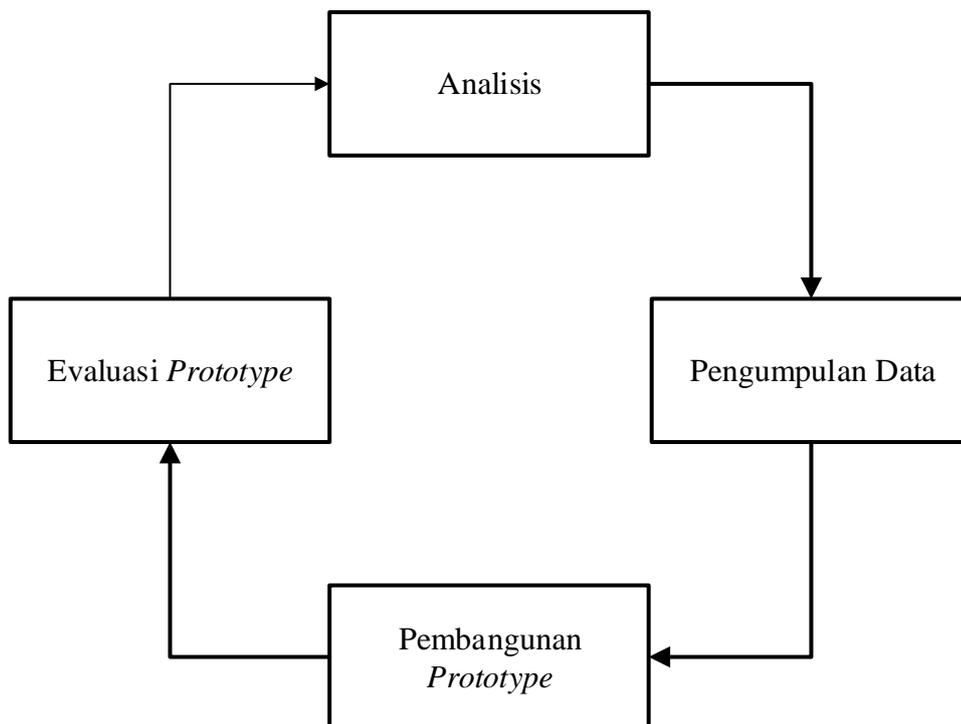
Pembangunan sistem *typo checking* menggunakan algoritma *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch* merupakan tahapan dimana peneliti melakukan penerapan analisis yang sebelumnya telah dilakukan lalu menguji coba terhadap sistem yang telah dibuat dan menganalisa performansi dari algoritma yang digunakan.

6. Pengujian

Pengujian merupakan tahapan dimana peneliti melakukan uji sistem atas apa yang sebelumnya telah dibuat.

1.6. Metode Pembangunan Sistem

Metode pembangunan perangkat lunak yang digunakan adalah metode *prototype*, karena pada pembangunan penelitian pendeteksian pengetikan kesalahan kata didasarkan pada ide pengembangan awal, diuji keberhasilannya, dan menyempurnakannya dengan kembali ke pembangunan. Tahapan dari metode *prototype* dapat dilihat pada Gambar 1.2.



Gambar 1.2 Model Prototype [8]

1. Analisis

Analisis yang dilakukan yaitu mulai penerimaan input teks dokumen hingga melakukan proses pendeteksian kesalahan pengetikan, analisis penentuan rekomendasi kata dengan *levensthein*, *smith waterman* dan *needleman Wunsch* dan pemilihan perbaikan kata.

2. Pengumpulan Data

Pada tahap ini peneliti mengumpulkan berupa file teks Berbahasa Indonesia untuk data masukan sistem.

3. Pembanguna *Prototype*

Merupakan tahap pengimplementasian dari proses analisis dan kebutuhan sistem yang sudah didapatkan. Pada tahap ini pengembang mencoba mengimplementasikan metode *levensthein*, *smith waterman* dan *needleman wunsch* pada pendeteksian kesalahan pengetikan kata ke dalam logika-logika program.

4. Evaluasi *Prototype*

Program akan diuji, dimana uji coba dilakukan untuk mengetahui kekurangan pada program. Jika masih ada kekurangan, maka prototype direvisi dengan tahapan-tahapan yang sebelumnya telah dilakukan. Proses ini akan terus berulang sampai prototype yang dihasilkan mendekati harapan yang diinginkan.

1.7. Sistematika Penulisan

Sistematika penulisan laporan akhir penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tugas akhir ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian serta sistematika penulisan yang dimaksudkan agar dapat memberikan gambaran tentang urutan pemahaman dalam menyajikan laporan ini.

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai landasan teori yang digunakan untuk menganalisis masalah dan teori yang dipakai dalam data penelitian yaitu teori

mengenai algoritma *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch* dalam *typo checking*.

BAB 3 ANALISIS DAN PERANCANGAN

Pada bab ini dibahas mengenai kebutuhan algoritma *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch* dalam merekomendasikan kata yang salah tulis. Bab ini juga menjelaskan bagaimana algoritma tersebut dapat melakukan pemilihan kata. Selain itu, perancangan UML, Use Case Diagram, Activity Diagram, Class Diagram dalam perancangan antarmuka dari sistem yang dibuat akan dibahas pada bab ini.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini akan dijelaskan tahapan-tahapan implementasi dari kebutuhan algoritma dan pengujian berdasarkan analisis yang sudah dilakukan sebelumnya. Membandingkan algoritma *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch* mulai dari input dokumen, pencarian kata dengan cara mencocokkan dengan kata yang ada pada kamus kata. Pada saat setelah kata di cari kata tersebut akan diperbaiki dengan cara membandingkannya dengan kata yang ada pada kamus. Hasil akhir dari pengecekan sistem akan memberikan output berapa banyak saran perbaikan kata, pemeriksaan dan perbaikan struktur kalimat.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil Perbandingan *Levenshtein*, *Smith-Waterman*, dan *Needleman-Wunsch* Dalam Typo Checking serta saran untuk pengembangan media pembelajaran kedepannya.