

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Corpus merupakan salah satu elemen penting di dalam banyak bidang *Natural Language Processing* (NLP). Dalam banyak bidang NLP, seperti *part-of-speech (POS) tag*, *named entity recognition (NER)*, *co-reference*, dan *word disambiguation*, *corpus* yang telah dianotasi atau disebut dengan *annotated corpus* digunakan sebagai *data training*. *Data training* yang berkualitas tentu akan mempengaruhi hasil pelatihan dan evaluasi dari kebanyakan *tools* di bidang NLP. Namun, jumlah *annotated corpus* dalam bahasa Indonesia masih sedikit [1], sehingga penelitian-penelitian di bidang NLP dalam bahasa Indonesia belum bisa optimal.

Salah satu bidang dalam NLP adalah *POS tag*. *POS tag* memiliki karakteristik dimana setiap kata atau *token* diberi *tag* yang sesuai dengan kelasnya. Aplikasi anotasi kelas kata (*POS tag*) untuk teks berbahasa Indonesia sebelumnya belum pernah dibuat. Untuk teks berbahasa lain, misalnya Bahasa Inggris sudah pernah dibuat [2]. Namun, *tagset* untuk Bahasa Inggris dan Bahasa Indonesia berbeda. Tidak semua *tagset* yang ada dalam Bahasa Inggris sesuai dengan tata Bahasa Indonesia. Jika dilihat dari karakteristik *POS tag*, dimana setiap kata atau tokennya diberi *tag* yang sesuai dengan kelasnya, serta jumlah *tagset* yang tidak sedikit, aplikasi tersebut juga dirasa tidak cukup baik dalam menangani anotasi kelas kata.

Berdasarkan uraian yang telah dijelaskan, maka diperlukan sebuah aplikasi untuk mempermudah proses anotasi kelas kata untuk teks berbahasa Indonesia. Maka, pada penelitian ini akan dilakukan pembangunan aplikasi anotasi kelas kata untuk teks berbahasa Indonesia.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, maka identifikasi masalah dalam penelitian ini adalah belum adanya aplikasi anotasi kelas kata untuk Bahasa Indonesia, sehingga diperlukan aplikasi yang dapat mempermudah proses anotasi kelas kata untuk teks berbahasa Indonesia.

1.3 Maksud dan Tujuan

Berdasarkan masalah yang telah diuraikan pada bagian latar belakang dan identifikasi masalah, maka maksud dari penelitian ini adalah membangun aplikasi anotasi kelas kata untuk teks berbahasa Indonesia. Sedangkan tujuan yang ingin dicapai dari penelitian ini adalah menyediakan aplikasi anotasi kelas kata yang efektif untuk teks berbahasa Indonesia.

1.4 Batasan Masalah

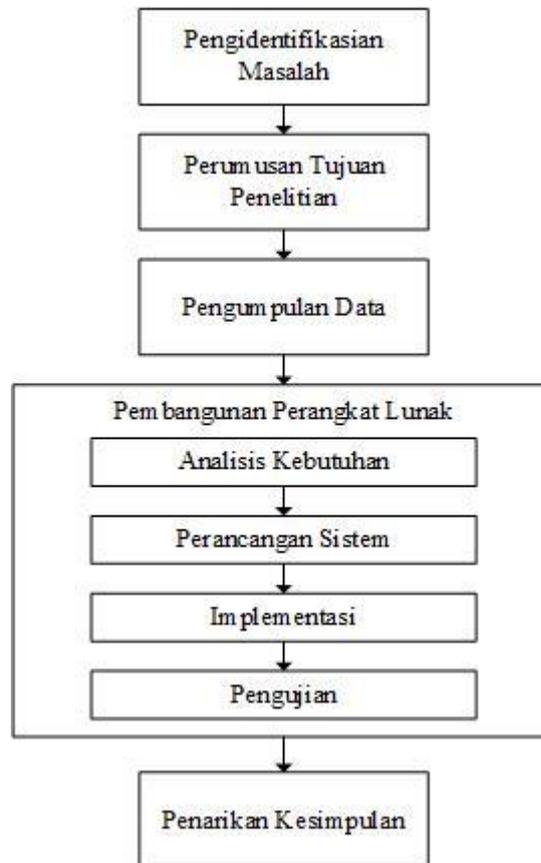
Agar permasalahan yang ditinjau tidak meluas dari maksud dan tujuan awal yang ingin dicapai, maka terdapat batasan-batasan masalah sebagai berikut.

1. Data masukan berupa berkas teks dalam format .txt, berisi artikel yang diambil dari situs berita dalam Bahasa Indonesia (SINDO News).
2. Proses yang terdapat di dalam proses anotasi yaitu *token matching* dan *multi-token selection*.
3. Data keluaran yang dihasilkan oleh aplikasi berupa *annotated corpus* dalam berkas teks dengan format .txt.

1.5 Metodologi Penelitian

Metodologi penelitian merupakan tahap-tahap yang telah ditentukan dalam melakukan sebuah penelitian yang berguna sebagai pedoman dalam melakukan proses penelitian agar penelitian yang dilakukan dapat berjalan dengan baik dan sistematis. Metodologi penelitian yang digunakan dalam penelitian ini adalah metode deskriptif [3]. Metode deskriptif digunakan karena pada penelitian ini karakteristik subjek atau objek penelitian digambarkan secara sistematis. Metodologi penelitian dibagi ke dalam dua jenis, yaitu metode pengumpulan data

dan metode pengembangan perangkat lunak. Adapun skema dalam penelitian ini yang dapat dilihat di Gambar 1.1.



Gambar 1.1 Skema Penelitian

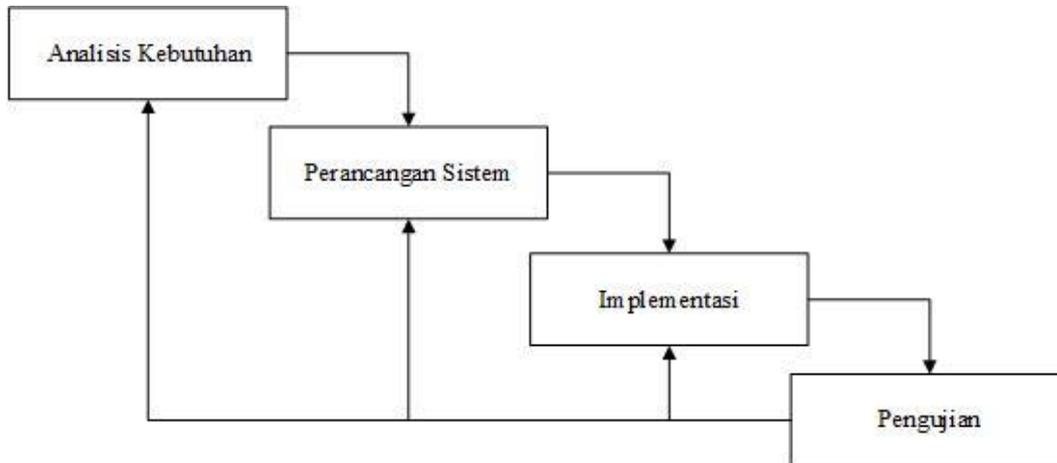
1.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah studi literatur, dengan mempelajari dokumen-dokumen tertulis baik yang tercetak maupun elektronik, seperti buku, e-book, jurnal, paper, dan sumber-sumber lainnya yang berhubungan dengan *POS tag*.

1.5.2 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan adalah metode *waterfall sommerville*. Metode *waterfall* atau *linear sequential model* merupakan model klasik yang bersifat sistematis, berurutan dalam membangun *software* [4]. Alur proses metode *waterfall*. Metode *waterfall sommerville* digunakan pada penelitian ini karena setiap tahap dilakukan secara runut, namun memungkinkan

untuk kembali ke tahap sebelumnya apabila terdapat kesalahan atau perbaikan pada tahap yang sedang dijalankan. Tahap-tahap pada metode *waterfall sommerville* dapat dilihat di Gambar 1.2 [5].



Gambar 1.2 Alur Tahapan Metode *Waterfall Sommerville*

Rincian tahap pada metode *waterfall sommerville* adalah sebagai berikut.

1. Analisis Kebutuhan

Tahap ini menentukan permasalahan, kebutuhan, dan tujuan yang ingin dicapai untuk kemudian didefinisikan secara rinci dan berfungsi sebagai spesifikasi sistem.

2. Perancangan Sistem

Pada tahap ini, arsitektur sistem dibentuk berdasarkan spesifikasi yang telah didefinisikan di tahap sebelumnya.

3. Implementasi

Tahap ini merupakan realisasi dari perancangan sistem yang telah dibuat pada tahap sebelumnya ke dalam sebuah program.

4. Pengujian

Pada tahap ini, program diuji untuk memastikan sistem sudah memenuhi persyaratan yang telah ditetapkan.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini dibuat dengan tujuan untuk memberikan gambaran umum dari penelitian yang akan dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut.

BAB 1 Pendahuluan

Bab ini menguraikan tentang latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian yang digunakan, dan sistematika penulisan pada penelitian yang dilakukan.

BAB 2 Landasan Teori

Bab ini menguraikan tentang teori atau konsep yang berkaitan dengan pembangunan aplikasi anotasi kelas kata untuk teks dalam bahasa Indonesia, baik yang berasal dari buku, jurnal, maupun artikel. Teori atau konsep tersebut terdiri dari *Natural Language Processing*, *POS tag*, *tagset*, serta metode dan alat-alat (*tools*) pendukung dalam pembangunan aplikasi.

BAB 3 Analisis dan Perancangan

Bab ini menguraikan tentang analisis yang memaparkan proses identifikasi masalah pada aplikasi yang dibangun, analisis kebutuhan non-fungsional dan fungsional, serta perancangan antarmuka yang menggambarkan rancangan aplikasi yang di bangun.

BAB 4 Implementasi dan Pengujian

Bab ini menguraikan tentang hasil dari analisis dan perancangan yang telah dibuat sebelumnya, yaitu aplikasi anotasi kelas kata untuk teks dalam Bahasa Indonesia. Aplikasi kemudian diuji untuk mengetahui tujuan awal penelitian dapat tercapai atau tidak.

BAB 5 Kesimpulan dan Saran

Bab ini menguraikan tentang hasil yang didapat dari penelitian yang telah dilakukan, serta masukan untuk pengembangan lebih lanjut agar penelitian yang akan datang dapat memberikan hasil yang lebih baik.

