

1. Pendahuluan

1.1 Latar Belakang

Bahasa merupakan alat berkomunikasi antar manusia untuk menyampaikan berbagai hal ke sesama. Bahasa untuk setiap negara tentunya memiliki pola dan struktur tersendiri. Sebagai contohnya untuk Bahasa Indonesia baku (berdasarkan Ejaan yang Disempurnakan) dimana jenis kata terbagi dalam 7 jenis utama, yaitu kata kerja, kata benda, kata sifat, kata keterangan, kata ganti, kata bilangan, dan kata tugas. Tiap jenis kata tersebut memiliki makna dan perannya masing-masing. Seiring dengan kemajuan teknologi banyak pihak yang ingin membuat mesin yang dapat memahami dan berkomunikasi bahasa manusia yang disebut sebagai *Natural Language Processing* (NLP).

Salah satu bagian penting dalam NLP adalah *Part-of-Speech* (POS) *Tagging*. POS *Tagging* dapat disebut juga *grammatical tagging* merupakan suatu proses memberikan kelas kepada suatu kata dalam sebuah kalimat atau *text*. POS *Tagging* sebagai bagian dari NLP memiliki banyak aplikasi diantaranya untuk *machine translation*, *parsing*, *word sense disambiguation*, dan *question answering*.

Sebenarnya POS *Tagging* sudah cukup lama dikenal namun implementasinya untuk Bahasa Indonesia masih jarang ditemukan. POS *Tagging* Bahasa Indonesia masih didominasi oleh pembelajaran pola berdasarkan aturan-aturan dari kalimat yang sudah ditentukan jenis katanya oleh ahli bahasa. Hal tentunya memakan banyak waktu, dan membutuhkan campur tangan ahli bahasa.

Ketergantungan akan ahli bahasa ini dapat digantikan dengan cara mengenali jenis kata pada kalimat dalam suatu corpus melalui teknik *clustering* kata. Terdapat berbagai macam metode *clustering* salah satunya adalah *graph clustering*. Dalam *graph clustering* pun memiliki banyak algoritma penerapan Chinese Whisper merupakan salah satunya. *Chinese Whisper* merupakan algoritma *randomized graph-clustering* yang telah terbukti berguna dalam aplikasinya untuk NLP[1, 2]. Hasil *clustering* kata ini yang kemudian akan digunakan sebagai landasan pembangunan model Viterbi dalam penentuan jenis kata pada sebuah kalimat.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, berikut adalah rumusan masalah yang dapat dirumuskan:

- a. Bagaimana mengimplementasikan pengelompokkan dan pengenalan jenis kata (*POS Tagging*) dari artikel berbahasa Indonesia menggunakan metode *Graph Clustering* menggunakan algoritma *Chinese Whisper*?
- b. Berapa tingkat akurasi *POS Tagging* yang dapat dicapai?

1.3 Batasan Masalah

Batasan masalah dalam pengerjaan Tugas Akhir ini adalah:

- a. Data teks yang akan digunakan berasal dari *corpus* yang sudah ada
- b. Keluaran sistem yang dihasilkan adalah kata yang telah ditentukan kelas katanya yang berbentuk angka *cluster*
- c. Penentuan jenis kata suatu *cluster* dilakukan secara manual oleh penulis berdasarkan jenis kata paling sering muncul
- d. Tingkat akurasi penggunaan metode akan digunakan dari data yang dihasilkan dengan *corpus* asli yang telah ada
- e. *Tag set* yang digunakan sebanyak 36 jenis *tag*

1.4 Tujuan

Tujuan yang ingin dicapai dalam penelitian Tugas Akhir ini adalah sebagai berikut:

- a. Merancang dan membangun aplikasi yang dapat mengelompokkan dan mengenal jenis kata dari sebuah artikel berbahasa Indonesia menggunakan metode *Graph Clustering* menggunakan algoritma *Chinese Whisper*
- b. Menganalisis tingkat akurasi hasil *POS Tagging*

1.5 Metodologi Penyelesaian Masalah

Tahapan-tahapan yang dilakukan dalam penyelesaian tugas akhir ini adalah sebagai berikut:

- a. Pengumpulan data dan studi literatur
Membaca dan mempelajari sumber tertulis seperti buku, artikel, internet, dan literatur yang berhubungan dengan metode *Chinese Whisper*, Viterbi, dan Part of Speech Tagging.
- b. Perancangan model dan sistem
Perancangan sistem menggunakan metode *Object Oriented*.
- c. Implementasi sistem
Sistem ini akan diimplementasikan menggunakan bahasa pemrograman Java.
- d. Pengujian sistem
Pengujian sistem dilakukan untuk mencari *bug* dan menguji keberhasilan implementasi sistem yang telah dibuat. Hasil pengujian sistem akan digunakan sebagai bahan analisis.
- e. Analisis hasil
Analisis sistem dilakukan untuk mengetahui tingkat akurasi dari sistem POS Tagging yang dihasilkan berdasarkan hasil implementasi dan pengujian sistem sesuai dengan rumusan masalah yang telah dipaparkan.
- f. Pembuatan laporan
Tahapan terakhir dari penelitian ini adalah pembuatan laporan, berfungsi sebagai dokumentasi terhadap proses dan hasil penelitian yang telah dilakukan.