I. PENDAHULUAN

I.1 Latar belakang masalah

Information Retrieval System merupakan sistem pengambilan informasi yang dapat diimplementasikan pada pencarian kata pada isi dan konteks dokumen. Dalam me*retrieve* berbagai jenis dokumen dilakukan *preprocessing* untuk mengambil informasi yang ada. *Preprocessing* sendiri terdiri dari *tokenization, Stoplist Removal,* dan *Stemming.*

Stemming merupakan suatu teknik untuk mentranformasi kata-kata dalam sebuah dokumen teks menjadi bentuk kata dasar[1]. Proses *stemming* berbeda dalam tiap bahasa karena pada setiap bahasa yang digunakan di berbagai negara memiliki aturan-aturan yang berbeda dalam penggunaan kata berimbuhan[4]. Bahasa Perancis memiliki perbedaan aturan penggunaan tata bahasa dengan bahasa Arab. Pada bahasa Indonesia terdapat komplekstias pada variasi imbuhan yang menjadi titik fokus pada pembentukan kata dasarnya. Algoritma Stemming yang digunakan pertama kali untuk menstemming bahasa Indonesia adalah Algoritma Nazief-Adriani (1996), mengacu pada algoritma Porter Stemmer yang digunakan pada bahasa Inggris. Algoritma stemming mengalami perkembangan untuk meminimalisir kekurangan-kekurangan yang ada, setelah Algoritma Nazief-Adriani selanjutnya ada algoritma Vega (2001), algoritma Arifin-Setiono (2002) dan algoritma Confix Stripping Stemmer (2007). Yang diangkat pada tugas akhir ini adalah algoritma Enhanced Confix Stripping Stemmer (2008), merupakan algoritma perbaikan dari Confix Stripping Stemmer.

Algoritma Enhanced Confix Stripping (ECS) Stemmer dapat digunakan untuk melakukan stemming pada dokumen teks bahasa Indonesia[3]. Algoritma *ECS Stemmer* memiliki beberapa kelemahan, diantarannya keterbatasan dalam me*nstemming* kata yang memiliki sisipan, lalu kekurangan mengenai overstemming. Oleh sebab itu, dalam Tugas Akhir ini, diajukan modifikasi perbaikan terhadap algoritma *ECS Stemmer* untuk mengatasi kelemahan tersebut sehingga mendapatkan tingkat akurasi yang lebih baik.

I.2 Perumusan masalah

Permasalahan-permasalahan yang akan diteliti pada tugas akhir ini antara lain:

Bagaimana menghasilkan tingkat akurasi yang lebih baik dari algoritma-algoritma *stemming* untuk bahasa Indonesia sebelumnya dengan memodifikasi algoritma Enhanced Confix Stripping Stemmer?

Adapun batasan-batasan masalah pada Tugas Akhir ini antara lain :

a. Teks yang digunakan sebagai dokumen uji merupakan novel yang menggunakan bahasa Indonesia yang baku.
b. Kamus yang menjadi acuan adalah KBBI.
c. Teks uji yang digunakan sudah dialihkan menjadi bertipe .txt
d. Sistem yang akan dibangun menggunakan bahasa pemrograman java.

I.3 Tujuan

Mengacu pada masalah-masalah diatas, tujuan Tugas Akhir ini adalah :

Menghasilkan tingkat akurasi yang lebih baik dari algoritma-algoritma *stemming* untuk bahasa Indonesia sebelumnya dengan memodifikasi algoritma Enhanced Confix Stripping Stemmer

I.4 Hipotesa

Dengan menambahkan skema baru pemotongan imbuhan dan aturan tambahan tabel pemotongan imbuhan pada Modifikasi Algoritma Enhanced Confix Stripping Stremmer untuk *stemming* pada teks bahasa Indonesia akan menghasilkan akurasi yang tinggi (rata-rata lebih dari 90% dari 4 novel dokumen uji) daripada Algoritma Enhanced Confix Stripping Stemmer murni tanpa modifikasi.

I.5 Metodologi Penyelesaian

Metodologi penyelesaian masalah yang akan dilakukan pada penelitian Tugas Akhir ini adalah sebagai berikut :

1. Studi literatur

   Tahap ini akan melakukan pencarian referensi dan materi yang ada, berupa paper, jurnal internasional dan buku. Memahami dan mempelajari referensi tersebut untuk menyelesaikan permasalahan dalam tugas akhir ini. Pencarian referensi meliputi studi pustaka tentang:

   a. Stemming
   b. Enhanced Confix Stripping Stremmer
   c. Aturan pemenggalan imbuhan bahasa Indonesia
   d. Algoritma Confix Stripping Stemmer

2. Pengumpulan data

   Mengumpulkan dataset aturan pemenggalan imbuhan pada teks bahasa Indonesia, mengumpulkan dokumen uji, stoplist dan kamus acuan berdasarkan KBBI.

3. Analisis dan perancangan sistem

   Melakukan analisis *stemming* pada teks bahasa Indonesia, merancang sistem menggunakan data yang sudah dikumpulkan.

4. Implementasi dan pembangunan sistem

   Melakukan pengimplementasian sistem, akan dituangkan kedalam sebuah program yang bisa melakukan proses *stemming* pada teks bahasa Indonesia secara akurat.

5. Pengujian sistem dan analisa hasil

   Melakukan pengujian terhadap program yang sudah dirancang sedemikian rupa agar mendapatkan hasil yang selanjutnya akan dianalisis.

6. Pengambilan kesimpulan dan penyusunan laporan

   Melakukan pengambilan kesimpulan dari hasil penelitian dan melakukan penyusunan laporan Tugas Akhir.

I.6 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini dibahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan, metodologi penelitian, dan sistematika penulisan Tugas Akhir ini.

BAB II LANDASAN TEORI

Pada bab ini dibahas mengenai teori-teori yang digunakan dalam penyusunan Tugas Akhir. Teori yang terdapat pada bab ini mencakup pengertian *Stemming*, *preprocessing*, algoritma Enhanced Confix Stripping Stemmer.

BAB III PERANCANGAN SISTEM

Pada bab ini dibahas mengenai langkah-langkah dalam mengidentifikasi *Stemming*, perancangan sistem (*user interface*).

BAB IV IMPLEMENTASI DAN ANALISIS

Pada bab ini dibahas mengenai implementasi modifikasi algoritma Enhanced Confix Stripping Stemmer, uji coba terhadap program yang telah dibuat, dan melakukan analisis terhadap hasil yang didapat dari implementasi.

BAB V PENUTUP

Pada bab ini berisi kesimpulan yang didperoleh dari pembuatan Tugas Akhir ini dan saran yang mungkin dapat berguna dalam penelitian selanjutnya.