

1. Pendahuluan

1.1 Latar Belakang

Salah satu bagian penting dari *Information retrieval* adalah proses *stemming*. *Stemming* adalah proses mereduksi kata berimbuhan menjadi bentuk umumnya. *Stemming* sangat berguna untuk proses *indexing* maupun *searching* di dalam *Information Retrieval*. Dengan proses *Stemming*, kata yang dimasukkan ke dalam *index* adalah dalam bentuk umum, sehingga dapat menghasilkan dokumen yang lebih relevan.

Dalam tugas akhir ini, akan dilakukan implementasi dan analisis perbandingan algoritma *stemming* Nazief & Adriani dengan Algoritma *stemming* Vega di dalam *Information Retrieval System*. Algoritma Nazief & Adriani mengeliminasi imbuhan secara satu per satu dengan menggunakan aturan tertentu, dimulai dari sufiks terlebih dahulu kemudian prefiks. Di setiap langkah eliminasi, akan dilakukan pengecekan apakah kata tersebut terdapat di dalam kamus. Kamus tersebut berisi kata-kata dasar di dalam Bahasa Indonesia.

Berbeda dengan algoritma Nazief & Adriani, Algoritma Vega tidak menggunakan kamus. Algoritma Vega menghilangkan imbuhan dengan aturan tertentu, dimulai dari prefiks terlebih dahulu kemudian sufiks. Jika ditemukan imbuhan, maka akan dilakukan penghapusan dengan aturan tertentu terhadap imbuhan tersebut tanpa dilakukan pengecekan dengan kamus.

Alasan pemilihan algoritma Vega adalah karena algoritma Nazief & Adriani memiliki waktu eksekusi yang cukup besar. Untuk itu, akan diteliti apakah algoritma Vega dapat menutupi kekurangan algoritma Nazief & Adriani di dalam segi waktu eksekusi. Selain itu, akan diteliti bagaimana performansi antara algoritma Nazief & Adriani dan algoritma Vega jika dilihat dengan parameter *Word Conflation Class*, dan *Index Compression Factor*. Kedua parameter tersebut merupakan parameter yang sering digunakan untuk mengukur proses *stemming*.

Untuk mengetahui hasil dari algoritma *stemming* Nazief & Adriani dan algoritma *stemming* Vega, akan diukur menggunakan parameter waktu, akurasi, *Word Conflation Class*, dan *Index Compression Factor*, dan dibuat *information*

retrieval system yang dapat memperlihatkan performansi yang dihasilkan kedua algoritma tersebut.

1.2 Perumusan masalah

Berdasarkan latar belakang diatas, permasalahan yang menjadi fokus di dalam tugas akhir ini adalah :

1. Bagaimana menerapkan algoritma *stemming* Nazief & Adriani dalam *information retrieval system*?
2. Bagaimana menerapkan algoritma *stemming* Vega dalam *information retrieval system*?
3. Bagaimana mengukur performansi algoritma *stemming* Nazief & Adriani dan algoritma *stemming* Vega dalam *information retrieval system*?

1.3 Batasan Masalah

Batasan masalah dalam tugas akhir ini adalah :

1. Dokumen yang digunakan adalah dokumen berbahasa Indonesia.
2. Dokumen disimpan secara *offline* dalam bentuk .txt, dengan jumlah tetap dan bukan dokumen yang selalu di-*update*.
3. Koleksi dokumen disimpan dalam format tertentu, dimana di dalam dokumen tersebut terdapat judul dan isi berita.
4. Algoritma *stemming* yang digunakan adalah algoritma Nazief & Adriani dan algoritma Vega.
5. Fokus tugas akhir terletak pada proses *stemming*.
6. Kamus yang digunakan terdiri dari 28.526 kata yang mengacu pada Kamus Besar Bahasa Indonesia.

1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah:

1. Mengimplementasikan algoritma *stemming* Nazief & Adriani dalam *information retrieval system*.
2. Mengimplementasikan algoritma *stemming* Vega dalam *information retrieval system*.

3. Mengukur performansi algoritma *stemming* Nazief & Adriani dan algoritma *stemming* Vega dengan parameter waktu, akurasi, *Word Conflation Class*, dan *Index Compression Factor*.
4. Mengukur hasil *Information Retrieval System* dengan memperhatikan hasil yang didapat, dan tingkat kemiripan query dengan dokumen yang diterima.

1.5 Metodologi Penyelesaian Masalah

Metodologi yang digunakan dalam penyelesaian masalah - masalah yang ada pada kasus tugas akhir ini adalah sebagai berikut:

1. Identifikasi Masalah dan Solusi

Memilih kasus dan merumuskan masalah yang dimiliki pada kasus tersebut serta merancang solusi berupa pengembangan sistem.

2. Studi *Literature*

Di dalam tahap ini dilakukan pendalaman materi melalui studi pustaka dan referensi dari berbagai sumber seperti buku, dan jurnal.

3. Pengumpulan Data

Di dalam tahap ini dilakukan pengumpulan data. Seluruh dokumen diambil dari situs berita online <http://www.kompas.com>.

4. Perancangan Sistem

Dilakukan analisis dan perancangan sistem untuk tahap implementasi.

5. Implementasi

Mengimplementasikan sistem dan komponen yang dirancang sebelumnya ke dalam bentuk aplikasi.

6. Pengujian dan Analisis Hasil

Dilakukan Pengujian dan Analisis terhadap hasil implementasi.

7. Pembuatan Laporan

Pada tahap ini disusun laporan tugas akhir.