

BAB I

PENDAHULUAN

1 Latar Belakang

Kebutuhan akan informasi sangat meningkat. Banyak informasi yang perlu diperoleh dengan cepat, dan relevan sesuai kebutuhan, tetapi untuk mendapatkan informasi yang relevan dan akurat sangatlah sulit. Berdasarkan hal tersebut, diperlukan suatu sistem untuk memenuhi kebutuhan informasi tersebut. Information retrieval adalah metode pencarian bahan (dokumen), termasuk proses menemukan dan mengambil informasi berbasis pengetahuan dan kumpulan dokumen. Information retrieval didefinisikan sebagai tindakan, metode dan prosedur untuk mengambil data yang disimpan, kemudian memberikan informasi tentang subjek yang diperlukan. Information retrieval digunakan untuk mengambil dokumen teks yang dibutuhkan oleh pengguna, dokumen teks tersebut merupakan dokumen digital. Sebagian besar mesin pencari konvensional bekerja berdasarkan query teks, yaitu pengguna memasukkan sejumlah kata sebagai query dan mesin pencari mencari dokumen yang terkait dengan query tersebut.

Di internet terdapat banyak sekali dokumen berbahasa Indonesia dalam bentuk arikel, dokumen,jurnal dan lain-lain. Afiks mengambil peran yang sangat penting dalam semua bahasa, termasuk bahasa Indonesia itu sendiri. Bahasa Indonesia memiliki afiks yang terdiri dari prefixes, suffixes, dan infixes. Sebuah kata dalam bahasa Indonesia dapat memiliki satu jenis afiks atau bahkan gabungan

dari semua jenis afiks, sebagai contoh kata cangkul akan berubah menjadi angkul bila digabungkan dengan imbuhan meny-, hal tersebut menyebabkan sulitnya pencocokan kata yang berhubungan. Dalam information retrieval, sebelum pencarian informasi dokumen teks dilakukan pengolahan kata yang dikenal dengan stemming, yaitu proses untuk mendapatkan kata dasar dengan menghapus afiks dari kata yang dicari.

Stemming kata memiliki tugas yang merupakan salah satu Teknik preprocessing dasar dalam NLP (Natural Language Processing). Stemming memiliki banyak fungsi seperti pengurangan kosakata, dengan menghasilkan kosakata yang mudah dikelola, membuat perhitungan menjadi lebih cepat dan efisien. Stemming juga dapat menangani salah ejaan dengan mereduksinya menjadi bentuk kata dasar. Dan juga stemming dapat meningkatkan kinerja tugas pemrosesan bahasa alami tertentu, seperti sentiment analysis, atau model machine learning dengan berfokus pada makna inti kata, bukan variasinya. Banyak algoritma yang dikembangkan untuk melakukan proses stemming bahasa, diantaranya ada algoritma Nazief dan Adriani dan algoritma Arifin Setiono.

Dalam penelitian ini, akan dilakukan analisis petbandingan algoritma stemming Nazief & Adriani dengan algoritma stemming Arifin Setiono untuk stemming Bahasa Indonesia. Dalam proses stemming bahasa Indonesia kata yang dihasilkan kadang ambigu atau membingungkan karena aturan morfologi bahasa. Agar dalam proses stemming kedua algoritma dapat mengurangi keambiguan atau meningkatkan akurasi maka akan dilakukan pengecekan apakah kata tersebut

terdapat di dalam kamus besar bahasa Indonesia. Dengan melewati proses validasi menggunakan kamus, diharapkan ketepatan dapat ditingkatkan.

2 Rumusan Masalah

Berdasarkan latar belakang maka dirumuskan permasalahan yang akan diuraikan solusinya yaitu :

1. Bagaimana menerapkan algoritma stemming Nazief & Adriani dalam melakukan stemming Bahasa Indonesia?
2. Bagaimana menerapkan algoritma stemming Arifin Setiono dalam melakukan stemming Bahasa Indonesia?
3. Bagaimana mengukur performansi keakuratan algoritma stemming Nazief & Adriani dengan algoritma stemming Arifin Setiono berdasarkan pada parameter ketepatan akurasi waktu, word conflation dan index compression factor dalam melakukan stemming Bahasa Indonesia?

3 Batasan Masalah

Batasan masalah yang akan digunakan untuk mempersempit masalah sebagai berikut:

1. Dokumen yang digunakan adalah dokumen berbahasa Indonesia.
2. Algoritma stemming yang digunakan adalah algoritma Nazief & Adriani dan algoritma Arifin Setiono.
3. Kamus yang digunakan terdiri dari 30.138 kata dasar yang mengacu pada Kamus Besar Bahasa Indonesia.
4. Fokus penelitian terletak pada proses stemming.

5. Bahasa pemrograman yang digunakan adalah PHP dan MySQL serta xampp sebagai server.

4 Tujuan Penelitian

Tujuan penelitian yang dilakukan adalah sebagai berikut:

1. Mengimplementasikan algoritma stemming Nazief&Adriani dalam melakukan stemming Bahasa Indonesia.
2. Mengimplementasikan algoritma stemming Arifin Setiono dalam melakukan stemming Bahasa Indonesia.
3. Mengukur performansi algoritma stemming Nazief&Adriani dan algoritma stemming Arifin Setiono dengan parameter akurasi waktu, word conflation, dan index compression factor.
4. Mengukur hasil informasi setelah melakukan proses stemming dari kedua algoritma yang digunakan dengan memperhatikan tingkat kemiripan query dengan dokumen yang diterima.

5 Manfaat Penelitian

Manfaat dari penelitian yang dilakukan adalah sebagai berikut:

1. Mengetahui dan memahami cara mengimplementasi algoritma stemming Nazief&Adriani dengan algoritma Arifin Setiono pada aplikasi berbasis website.
2. Mengetahui perbandingan proses keakuratan algoritma stemming Nazief&Adriani dengan algoritma stemming Arifin Setiono.

3. Sebagai referensi untuk melakukan penelitian lebih lanjut tentang Stemming bahasa Indonesia
4. Sebagai referensi penggunaan algoritma Nazief&Adriani dan algoritma Arifin Setiono untuk stemming Bahasa Indonesia.

1.6 Sistematika Penulisan

Penulisan skripsi ini ditulis secara sistematis yang terdiri dari beberapa sub bab. Berikut uraian sistematika penulisan.

1.6.1 BAB I : Pendahuluan

Pada bab ini akan dijelaskan tentang latar belakang, rumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

1.6.2 BAB II : Tinjauan Pustaka

Pada bab ini akan dijelaskan tentang Kajian pustaka, Dasar Teori, Kutipan kutipan dari penelitian sebelumnya.

1.6.3 BAB III : Metode Penelitian

Pada bab ini akan menganalisis tentang desain system dari algoritma Nazief&Adriani dan Algoritma Arifin Setiono yang telah dikembangkan sebelumnya.

1.6.4 BAB IV : Implementasi dan Pembahasan

Pada bab ini akan membahas tentang hasil analisis dan penerapan algoritma Nazief&Adriani dan algoritma Arifin Setiono dalam melakukan proses stemming.

1.6.5 BAB V : Penutup

Pada bab ini akan berisikan kesimpulan dan saran dari perumusan masalah yang disampaikan.

