

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kebutuhan terhadap informasi sangat meningkat. Banyak informasi dibutuhkan untuk dapat diperoleh dengan cepat, dan relevan sesuai kebutuhan. tetapi mengambil informasi yang relevan dan berarti sangat sulit. Berdasarkan hal tersebut dibutuhkan sebuah sistem untuk memenuhi kebutuhan informasi. Information retrieval merupakan metode mencari material (dokumen), meliputi proses mencari dan mengambil informasi berbasis pengetahuan dari kumpulan dokumen. Information retrieval didefinisikan sebagai tindakan, metode, dan prosedur untuk menemukan kembali data yang tersimpan, kemudian menyediakan informasi mengenai subyek yang dibutuhkan. Information retrieval digunakan untuk menemukan kembali dokumen teks yang dibutuhkan oleh pengguna, dokumen teks tersebut berupa dokumen yang berbentuk digital. Sebagian besar mesin pencari konvensional bekerja berdasarkan text queries, yaitu pengguna memasukkan sejumlah kata sebagai query dan mesin pencari mencari dokumen yang berhubungan dengan query tersebut.

Di internet terdapat banyak sekali dokumen berbahasa Indonesia dalam bentuk artikel, jurnal, dan lain-lain. Bahasa Indonesia berbeda dari bahasa lain karena setiap bahasa memiliki karakteristik sendiri. Dalam Bahasa Indonesia, afiks mengambil peran penting dan sangat sensitif, karena sedikit perbedaan afiks dapat menyebabkan perbedaan makna yang jauh. Afiks terdiri dari prefixes, infixes, suffixes, dan confixes. Suatu kata dapat digabungkan dengan prefixes,

infixes, suffixes dan confixes, hal tersebut menyebabkan sulitnya pencocokan kata yang berhubungan. Beberapa penggabungan menyebabkan perubahan pada akar kata, sebagai contoh kata sapu akan berubah menjadi apu bila digabungkan dengan imbuhan meny-. Dalam information retrieval, sebelum pencarian informasi dokumen teks dilakukan prapengolahan terlebih dulu yang dikenal dengan stemming, yaitu proses untuk mendapatkan kata dasar dengan menghapus afiks dari kata yang dicari. Tugas stemming kata merupakan salah satu teknik preprocessing dasar dalam NLP (Natural Language Processing). Stemming adalah proses heuristik yang memotong ujung kata untuk menemukan kata dasar, dan seringkali mencakup pembuangan afiks. Stemming digunakan untuk mengubah kata yang mengandung afiks menjadi kata dasar Banyak algoritma yang dikembangkan untuk melakukan proses stemming Bahasa Indonesia, di antaranya algoritma Nazief dan Andriani, algoritma Porter.

Dalam penelitian ini, akan dilakukan analisis perbandingan algoritma stemming Nazief & Adriani dengan Algoritma stemming Porter untuk stemming Bahasa Indonesia. 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 Porter merupakan algoritma yang paling banyak digunakan untuk stemming bahasa Inggris, yang mana cara kerjanya Algoritma stemming Porter mencari kata dasar dengan menghilangkan afiks/imbuhan, kata yang dihasilkan kadang ambigu atau membingungkan karena aturan morfologi bahasa

Indonesia yang tidak konsisten. Agar dalam proses stemming algoritma Porter dapat mengurangi keambiguan atau meningkatkan akurasi maka diusulkan penggunaan kamus kata dasar. Setiap kata yang telah melewati proses pemotongan imbuhan dibandingkan dengan kata yang tersimpan dalam kamus kata dasar. Dengan melewati proses validasi menggunakan kamus, diharapkan dapat meningkatkan ketepatan.

### **1.2 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 Porter dalam melakukan stemming Bahasa Indonesia?
3. Bagaimana mengukur performansi keakuratan algoritma stemming Nazief & Adriani dengan algoritma stemming Porter berdasarkan pada parameter ketepatan akurasi waktu, word conflation class dan index compression factor dalam melakukan stemming Bahasa Indonesia?

### **1.3 Batasan Masalah**

Batasan masalah yang akan digunakan untuk mempersempit persembahan sebagai berikut :

1. Dokumen yang digunakan adalah dokumen berbahasa Indonesia.
2. Algoritma stemming yang digunakan adalah algoritma Nazief&Adiani dan algoritma Porter.

3. Focus penelitian terletak pada proses stemming.
4. Kamus yang digunakan terdiri dari 28.526 kata yang mengacu pada Kamus Besar Bahasa Indonesia.
5. Menggunakan Bahasa pemrograman PHP dan MySQL serta xampp sebagai server

#### **1.4 Maksud dan Tujuan Penelitian**

Tujuan penelitian yang dilakukan adalah sebagai berikut :

1. Mengimplementasikan algoritma stemming Nazief&Adriani dalam melakukan stemming Bahasa Indonesia.
2. Mengimplementasikan algoritma stemming Porter dalam melakukan stemming Bahasa Indonesia.
3. Mengukur performansi algoritma stemming Nazief&Adriani dan algoritma stemming Porter dengan parameter akurasi waktu, word conflation class, 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.

#### **1.5 Manfaat Penelitian**

Manfaat dari penelitian yang dilakukan adalah sebagai berikut :

1. Mengetahui dan memahami cara mengimplementasi algoritma stemming Nazief & Adriani dengan algoritma stemming Porter pada aplikasi berbasis website.

2. Mengetahui perbandingan proses keakuratan algoritma stemming Nazief & Adriani dengan algoritma stemming Porter.
3. Sebagai referensi untuk melakukan penelitian lebih lanjut tentang Stemming khususnya dengan pendekatan morfologi Bahasa Indonesia.
4. Sebagai referensi penggunaan algoritma Porter untuk stemming Bahasa Indonesia.

## **1.6 Metode Penelitian**

Penelitian ini menggunakan metode penelitian Research and Development (penelitian dan pengembangan). Metode ini adalah metode penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut. Hasil penelitian ini diharapkan nantinya digunakan sebagai masukan untuk berbagai jenis penelitian penggalan text khususnya proses stemming perbandingan metode stemmer pada tahap pra-proses penggalan teks, sedangkan secara praktik pada aplikasinya nantinya dapat digunakan untuk orang yang melakukan penelitian penggalan text yang mengalami masalah seperti Understemming, Unchange, spelling exption, dll. Langkah-langkah yang dilakukan peneliti adalah sebagai berikut:

### **1.6.1 Metode Pengumpulan Data**

Pengumpulan data yang dilakukan untuk memperoleh data-data yang dilakukan dalam penelitian yaitu:

1. Metode study literatur

Mempelajari data-data dari jurnal, buku dan website yang berkaitan dengan algoritma yang akan diteliti. Tujuan dari metode ini adalah untuk memperkuat permasalahan dan juga sebagai dasar teori dalam melakukan penelitian

## 2. Metode study pustaka

Mempelajari data-data dan arsip yang sudah ada sebagai acuan atau bahan referensi dalam memperoleh data informasi yang dibutuhkan.

### 1.6.2 Metode Analisis

Metode analisis yang digunakan dalam penelitian ini yaitu analisis kebutuhan system yang meliputi kebutuhan fungsional dan non-fungsional dan analisis kelayakan system yang meliputi kelayakan teknologi, operasional dan hukum.

### 1.6.3 Metode Perancangan

Metode perancangan yang dilakukan yaitu membuat prototype terlebih dahulu dengan melakukan simulasi desain kebutuhan system seperti interface dan flowchart.

### 1.6.4 Metode Pengembangan

Pada tahap metode pengembangan akan dimulai membuat kode program berdasarkan desain yang telah dibuat sebelumnya.

### 1.6.5 Metode Testing

Pada tahap ini dilakukan perbandingan akurasi dan kecepatan pada algoritma Nazief&Adriani dan algoritma Porter.

## **1.7 Sistematika Penulisan**

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

### **1.7.1 BAB I : Pendahuluan**

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

### **1.7.2 BAB II : Landasan Teori**

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

### **1.7.3 BAB III : Analisis dan Perancangan**

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

### **1.7.4 BAB VI : Implementasi dan Pembahasan**

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

### **1.7.5 BAB V : Penutup**

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