

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Pencarian informasi berupa dokumen teks atau yang dikenal dengan istilah *Information Retrieval (IR)* merupakan proses pemisahan dokumen-dokumen yang dianggap relevan dari sekumpulan dokumen yang tersedia (Agusta, 2009). Salah satu bagian penting dari *Information Retrieval* adalah proses *stemming*. *Stemming* adalah proses mereduksi kata berimbuhan menjadi kata dasar. *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 (Rozi dkk, 2013). Metode *stemming* adalah salah satu cara yang digunakan untuk mengubah kata untuk menemukan akar kata dengan menerapkan aturan morfologi bahasa yang baik dan benar. Proses *stemming* dilakukan dengan menghilangkan semua imbuhan (*affiks*) baik yang terdiri dari awalan (*prefiks*) sisipan (*infiks*) maupun akhiran (*suffiks*) dan kombinasi awalan dan akhiran (*konfiks*) (Agusta, 2009).

Algoritma-algoritma *stemming* memiliki kelebihan dan kekurangannya masing-masing. Terdapat penelitian sebelumnya mengenai Perbandingan Algoritma Stemming Porter dan Algoritma Stemming Adriani Nazief Untuk Stemming Dokumen Teks Bahasa Indonesia yang menganalisis perbandingan pada dokumen berbahasa Indonesia. Berdasarkan hasil penelitian tersebut, disimpulkan bahwa algoritma porter lebih baik dalam hal kecepatan waktu namun memiliki kelemahan dalam hal keakuratan (Agusta, 2009). Penelitian lainnya menyebutkan bahwa algoritma Arifin Setiono digunakan karena memiliki kelebihan dalam hal mengatasi *Overstemming* yaitu jika kata tidak ditemukan setelah penghapusan maka algoritma ini kemudian mencoba untuk

mengembalikan semua kombinasi yang dihapus untuk mendapatkan kata yang valid (Asian, 2007).

Dalam penelitian ini akan dilakukan analisis performansi pada dokumen teks dengan menggunakan metode *stemming* dengan membandingkan dari tiga algoritma *stemming* Nazief & Adriani, Porter dan Arifin setiono yang nantinya akan diterapkan pada dokumen berbahasa Indonesia. Parameter yang akan diuji yaitu kecepatan dan ketepatan dari ketiga algoritma yang berpengaruh pada presentasi algoritma yang di implementasikan.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang masalah yang diuraikan, maka dapat dirumuskan masalah dalam penelitian ini yaitu:

1. Bagaimana performansi algoritma Nazief & Adriani dalam *stemming* teks berbahasa Indonesia.
2. Bagaimana performansi algoritma Porter dalam *stemming* teks berbahasa Indonesia.
3. Bagaimana performansi algoritma Arifin Setiono dalam *stemming* teks berbahasa Indonesia.
4. Bagaimana hasil perbandingan algoritma Nazief & Adriani, Porter dan algoritma Arifin Setiono untuk proses *stemming* teks berbahasa Indonesia.

### **1.3. Batasan Masalah**

Batasan masalah dari penelitian ini yaitu:

1. Dokumen yang digunakan adalah dokumen berbahasa Indonesia.
2. Parameter yang akan di hasilkan pada aplikasi ini adalah kecepatan dan ketepatan dari ketiga algoritma.
3. Kamus sebagai pembanding kata yang di *stemming* berupa kata dasar yang sesuai dengan Kamus Besar Bahasa Indonesia.

4. Tidak menggunakan tahap *filtering*.

#### **1.4. Tujuan Penelitian**

Tujuan dari tugas akhir ini adalah:

1. Mengetahui performansi berupa kecepatan dan ketepatan dari algoritma Nazief & Adriani untuk *stemming* teks berbahasa Indonesia.
2. Mengetahui performansi berupa kecepatan dan ketepatan dari algoritma Porter untuk *stemming* teks berbahasa Indonesia.
3. Mengetahui performansi berupa kecepatan dan ketepatan dari algoritma Arifin Setiono untuk *stemming* teks berbahasa Indonesia.
4. Membandingkan kecepatan dan ketepatan algoritma Nazief & Adriani, Porter dan algoritma Arifin Setiono untuk *stemming* teks berbahasa Indonesia.

#### **1.5. Manfaat Penelitian**

Manfaat dari penelitian ini adalah:

1. Mengetahui performansi berupa kecepatan dan ketepatan dari algoritma Nazief & Adriani, Porter dan algoritma Arifin Setiono dengan metode *stemming* yang diimplementasikan.
2. Sebagai dasar penelitian lain yang menggunakan *stemming*.