

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Bahasa Indonesia merupakan bahasa utama yang digunakan masyarakat Indonesia untuk berkomunikasi, baik secara lisan maupun tertulis. Dalam kehidupan sehari-hari mungkin tidak terlalu diperhatikan penulisan kata yang benar, namun dalam penggunaan surat resmi kata dalam Bahasa Indonesia menjadi sangat penting. Selain untuk berkomunikasi, penggunaan Bahasa Indonesia yang baik dan benar juga sangat diperlukan dalam penulisan-penulisan dokumen formal, jurnal, laporan dan sebagainya. Penggunaan bahasa yang benar menjadi unsur penting dalam Bahasa Indonesia, kamus Besar Bahasa Indonesia menjadi acuan utama dalam penggunaan kata baku Berbahasa Indonesia.

Pencarian informasi berupa dokumen teks atau yang dikenal dengan istilah *Information retrieval* (IR) merupakan proses pemisah dokumen-dokumen yang dianggap relevan dari sekumpulan dokumen yang tersedia. Bertambahnya jumlah dokumen teks yang dapat diakses diinternet dapat diikuti dengan meningkatnya kebutuhan pengguna akan perangkat pencarian informasi yang efektif dan efisien. Efektif berarti user mendapatkan dokumen yang relevan dengan *query* yang diinputkan. Efisien berarti waktu pencarian yang sesingkat-singkatnya.

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa IR (*Information Retrieval*) dengan cara mentransformasi kata-kata dalam sebuah teks ke kata dasarnya. Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma *stemming* untuk bahasa lainnya. Proses *stemming* pada teks Bahasa Indonesia lebih rumit karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan *root word* dari sebuah kata. Beberapa algoritma *stemming* bahasa indonesia telah dikembangkan sebelumnya. Penggunaan algoritma *stemming* yang sesuai mempengaruhi performa IR.

Algoritma *stemming* memiliki kelebihan dan kekurangannya. Efektifitas algoritma *stemming* dapat diukur berdasarkan beberapa parameter, seperti

kecepatan proses dan memiliki kelemahan dalam hal keakuratan, dan kesalahan *stemming*. Hasil dari *stemming* disebut token. Salah satu keuntungan menggunakan *stemming* dalam pengembangan sistem IR adalah efisiensi dan index file yang sudah terkomputasi.

Dalam Bahasa Indonesia ada beberapa penelitian yang dilakukan yang berhubungan dengan *stemming* Bahasa Indonesia. Semua penelitian yang telah dilakukan mengubah bentuk kata menjadi kata dasar dengan berdasarkan morfologi dari suatu kata. Penelitian sebelumnya yang dilakukan oleh Baskoro (2013) dari Ilmu Komputer Universitas Gadjah Mada dengan melihat struktur dari satu kata. Penelitian yang dilakukan oleh Baskoro lebih dulu melihat akhiran partikel dan kata ganti milik yang merupakan aturan struktur morfologi Bahasa Indonesia Infleksional kata ganti dan Derivasional sufik dibandingkan melihat awalan dari suatu kata, sedangkan masih banyak lagi aturan-aturan dari morfologi Bahasa Indonesia dari kedua Struktur tersebut. Algoritma ini menjadi tidak efisien karena diperlukan masing-masing 2 tabel di *database* untuk kata berakhiran partikel dan kata ganti milik. Selain itu kekurangan dari penelitian Baskoro adalah Algoritma ini tidak memperhitungkan kata dasar yang berawalan huruf “k”, “t”, “s”, “p” dengan benar. Dalam Bahasa Indonesia kata yang diawali dengan huruf “k”, “t”, “s”, “p” jika diberi awalan “me-“, atau “pe-“ maka huruf depannya akan melebur sebagai contoh : “mengacau” menjadi “me”+”kacau”. Jika kata “mengacau” dicari kata dasarnya dengan menggunakan *stemming porter* maka kata dasar yang muncul adalah “acau”. Kata ini tentu tidak bisa disamakan dengan kata “mengantar” menjadi “me” + “antar” walaupun secara struktur sama.

Hal ini tidak tepat jika digunakan untuk mencari kata dasar yang secara terstruktur. Metode *stemming* dengan menggunakan morfologi suatu kata memiliki beberapa kekurangan seperti tidak tepat menghilangkan awalan pada kata dasar yang berawalan huruf “k”, “t”, “s”, “p” serta tidak tepat dalam menghilangkan akhiran terutama untuk akhiran “kan” dan “an”. Untuk menyelesaikan masalah ini, penelitian ini menawarkan penggunaan 2 tingkat morfologi pada kata dasar berawalan huruf “k”, “t”, “s”, “p” serta menggunakan aturan kombinasi awalan dan akhiran, untuk menghilangkan akhiran pada suatu

kata seperti awalan “di” hanya boleh dipasangkan dengan akhiran “kan” dan tidak boleh dengan akhiran “an”. Hasil dari penelitian ini adalah *stemming* yang memiliki tingkat akurasi 95,5%, lebih baik dibandingkan *steeming* sebelumnya yang menggunakan algoritma berdasarkan morfologi satu kata. *Stemming* sebelumnya yang menggunakan algoritma berdasarkan morfologi suatu kata memiliki tingkat akurasi 82,5%.

Berdasarkan penelitian yang sudah ada, penulis mencoba melakukan penelitian bagaimana menentukan kata dasar pada teks bahasa indonesia dengan menggunakan “**peningkatan algoritma *stemming porter* bahasa indonesia berdasarkan metode morfologi dengan mengaplikasikan 2 tingkat morfologi dan aturan kombinasi awalan dan akhiran**”. Penulis ingin melakukan analisis akurasi algoritma “stemming porter dengan mengaplikasikan morfologi dan aturan kombinasi awalan dan akhiran” pada teks bahasa indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan diatas, terdapat beberapa permasalahan yang akan diangkat dalam penelitian ini antara lain:

1. Bagaimana menentukan kata dasar serta menghilangkan partikel awalan akhiran dalam sebuah dokumen bahasa indonesia menggunakan metode “peningkatan algoritma *stemming porter* Bahasa Indonesia berdasarkan metode morfologi dengan mengaplikasikan aturan kombinasi awalan akhiran”.
2. Bagaimana meningkatkan akurasi metode “peningkatan algoritma *stemming porter* bahasa indonesia berdasarkan metode morfologi dengan mengaplikasikan aturan kombinasi awalan akhiran”.

1.3 Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Penerapan algoritma pada sistem hanya memproses dokumen Berbahasa Indonesia. Dokumen yang akan digunakan adalah dokumen jurnal.

2. Parameter yang akan dihasilkan pada aplikasi ini adalah ketepatan dalam menentukan kata dasar.
3. Peningkatan pada penelitian ini terletak pada dua morfologi Bahasa Indonesia yaitu struktur infleksional dan struktur derivasional.
4. Aplikasi yang digunakan adalah aplikasi dengan menggunakan bahasa PHP (PHP: *hypertext preprocessor*).

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Mengolah dokumen Bahasa Indonesia untuk menentukan kata dasarnya.
2. Untuk mengetahui performa ketepatan dan akurasi dari metode “peningkatan algoritma *stemming porter* bahasa indonesia berdasarkan metode morfologi dengan mengaplikasikan aturan kombinasi awalan akhiran”.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini sebagai berikut:

1. Untuk perbaikan sistem pembelajaran mata pelajaran Bahasa Indonesia dengan memanfaatkan kemajuan teknologi.
2. Untuk memudahkan para guru, pelajar, maupun mahasiswa dalam proses belajar mengajar dalam menentukan kata dasar tanpa menggunakan Kamus Besar Bahasa Indonesia.
3. Memahami penerapan *steeming porter* pada dokumen Bahasa Indonesia.