

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Stemming merupakan sebuah cara untuk menghilangkan imbuhan seperti berupa awalan dan akhiran agar didapatkan kata dasar (*root word*) (Novitasari, 2016). Proses *stemming* memiliki aturan atau *algoritma* yang berbeda dalam menghilangkan imbuhan disetiap bahasanya, seperti bahasa inggris memiliki perbedaan aturan penggunaan tata bahasa dengan bahasa indonesia. *Algoritma Stemming* untuk bahasa indonesia yang telah dikembangkan diantaranya, *algoritma stemming Nazief & Adriani*, *algoritma stemming tala*, *algoritma stemming porter*. *Algoritma stemming Nazief & Adriani* mengembangkan berdasarkan aturan *morfologi* bahasa indonesia yaitu, mengelompokkan awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*), dan gabungan awalan akhiran (*confixes*). *Algoritma* ini menggunakan kamus kata dasar dan melakukan penyusunan kembali kata-kata, yang sudah mengalami proses *stemming (recoding)*.

Penelitian (Dwi Wahyudi, 2017) dalam kesimpulannya menyebutkan *Algoritma stemming Nazief & Adriani* mempunyai tingkat akurasi lebih baik daripada *algoritma stemming porter*, namun pada dasarnya masih terdapat kekurangan pada *algoritma stemming Nazief & Adriani* yaitu kurangnya aturan (*algoritma*) untuk menghilangkan imbuhan dan kesalahan urutan *algoritma*. Kurangnya aturan (*algoritma*) dan kesalahan urutan *algoritma* menimbulkan adanya pemenggalan imbuhan yang lebih (*overstemming*) atau pemenggalan imbuhan yang terlalu sedikit (*understemming*) sehingga kata dasar yang didapatkan tidak ada dalam kamus.

Penelitian yang akan dilakukan adalah memodifikasi metode *algoritma stemming Nazief & Adriani* untuk bahasa indonesia dengan metode *jaro winkler*. Metode *jaro winkler* pada penelitian ini digunakan untuk mencari kata dasar dengan mengukur kesamaan antara dua kata. Pada metode ini semakin tinggi nilai *jaro winkler* untuk dua buah kata, maka semakin mirip

kata tersebut. Nilai normalnya yaitu 0 untuk menandakan tidak ada kesamaan, dan 1 untuk menandakan adanya kesamaan (Ana Kurniawati & Rahman, 2010).

Pemilihan *algoritma stemming* Nazief & Adriani dipilih karena mempunyai tingkat akurasi lebih baik untuk mendapatkan kata dasar sedangkan metode *jaro winkler* dipilih karena metode ini dapat mengukur kesamaan antara dua kata. Pada permasalahan ini kata yang akan diukur yaitu kata yang dihasilkan oleh *algoritma stemming* Nazief & Adriani dengan kamus kata dasar, sehingga jika kedua *algoritma* ini digabungkan diharapkan dapat menemukan kata dasar yang tepat.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah :

“Bagaimana menangani pemenggalan imbuhan yang lebih (*overstemming*) atau pemenggalan imbuhan yang terlalu sedikit (*understemming*) ?”

1.3 Tujuan Penelitian

Adapun tujuan masalah pada penelitian ini adalah untuk mencari kata dasar pada kata yang mengalami pemenggalan imbuhan yang lebih (*overstemming*) atau pemenggalan imbuhan yang terlalu sedikit (*understemming*).

1.4 Manfaat penelitian

Manfaat penelitian ini berdasarkan latar belakang dan tujuan adalah untuk mendapatkan kata dasar yang sesuai dengan kamus, sehingga berguna mengurangi kesalahan dalam proses informasi *retrieval* dengan menambahkan metode *Jaro Winkler* pada *algoritma* Nazief & Adriani.

1.5 Batasan Masalah

Agar masalah yang akan dibahas tidak meluas, maka batasan masalah yang dibahas pada penelitian ini hanya difokuskan pada :

1. Dataset yang digunakan adalah kata dasar berbahasa indonesia, data diambil dari blog Putra Pandu Adikara pada postingan 17 november 2012.
2. *Query* yang diinputkan adalah sebuah kata.
3. Kata yang dilakukan *preprocessing* adalah kata berbahasa indonesia
4. Algoritma *Stemming* yang dimodifikasi adalah algoritma Nazief-Adriani.
5. Metode yang digunakan untuk meningkatkan algoritma *stemming* yaitu metode *Jaro Winkler*.
6. Hanya berlaku pada kata baku yang berkaidah EYD (Ejaan Yang Disempurnakan)
7. Sistem ini menggunakan bahasa pemrograman PHP dan *database* menggunakan MySQL.

1.6 Metodologi Penelitian

Metode yang digunakan dalam perancangan sistem antara lain:

1. Studi Literatur

Tahap ini akan melakukan pencarian referensi dengan cara mengumpulkan materi berupa paper, jurnal, dan buku. Memahami dan mempelajari referensi tersebut untuk menyelesaikan permasalahan sebagai penunjang penulisan skripsi.

2. Pengumpulan Data

Dalam proses pengumpulan data ini dilakukan pengumpulan dataset kata dasar bahasa indonesia.

3. Analisis Sistem

Analisis pada sistem dilakukan berdasarkan hasil observasi dari pengumpulan data. Analisa kebutuhan sistem dilakukan untuk menentukan fitur-fitur apa saja yang nantinya akan diterapkan pada sistem.

4. Implementasi sistem

Implementasi ini merupakan proses penerjemahan dari tahap perancangan ke dalam bentuk aplikasi dengan menggunakan bahasa pemrograman PHP dan *database* menggunakan MySQL.

5. Pengujian Sistem

Tahap Pengujian adalah tahap melakukan uji coba terhadap program yang dibangun untuk mengetahui sejauh mana kinerja sistem dan keakuratan metode yang diterapkan sehingga mampu menghasilkan informasi sesuai yang diharapkan.

6. Penulisan Laporan

Melakukan penyusunan laporan dari penelitian berdasarkan sistematika penulisan.

1.7 Sistematika Penulisan

Dalam sistematika penulisan ini terdapat pembahasan yang tersusun dalam beberapa kelompok sehingga mempermudah dalam memahami maksud dan tujuan dalam penelitian ini.

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang permasalahan, rumusan masalah yang akan diselesaikan, batasan masalah, tujuan penelitian, metode penelitian, sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas tentang teori-teori dasar yang berhubungan dengan permasalahan yang diambil, seperti penjelasan mengenai metode Nazief-Adriani *stemmer* dan metode *jaro winkler*.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini menjelaskan tentang penganalisaan kebutuhan dan perancangan dari sistem. Meliputi analisis sistem, analisis kebutuhan sistem, perancangan arsitektur, database, dan antarmuka sistem, serta langkah-langkah yang akan dilakukan untuk menyelesaikan permasalahan dan mencapai tujuan yang telah ditetapkan.

BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini menjelaskan tentang pengujian sistem secara umum maupun terperinci mengenai hasil penerapan sistem pada objek penelitian.

BAB V PENUTUP

Bab ini terdiri dari simpulan dan saran, yang berisi tentang simpulan hasil penelitian dan saran-saran yang dibutuhkan guna pengembangan sistem lebih lanjut.