

BAB I

PENDAHULUAN

1.1. Latar belakang

Teknik *stemming* adalah suatu teknik pencarian bentuk dasar dari suatu *term*. Yang dimaksud dengan *term* itu sendiri adalah tiap kata yang berada pada suatu dokumen teks. *Stemming* dilakukan pada saat pembuatan indeks dari suatu dokumen. Pembuatan indeks dilakukan karena suatu dokumen tidak dapat dikenali langsung oleh suatu sistem temu kembali informasi atau *information retrieval (IR) system*. Oleh karena itu, dokumen tersebut terlebih dahulu perlu dipetakan ke dalam suatu representasi dengan menggunakan teks yang berada di dalamnya. Teknik *stemming* diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau *form* yang berbeda karena mendapatkan imbuhan yang berbeda. [1]

Stemming umumnya digunakan pada pencarian teks, mesin terjemah, rangkuman dokumen dan klasifikasi teks. Di dalam proses *stemming*, *morfologi* adalah hal yang sangat diperhatikan dalam algoritmanya. *Morfologi* sendiri adalah suatu proses pembentukan kata, dan kata yang mengalami *morfologi* di dalam bahasa Indonesia adalah kata imbuhan, kata ulang dan kata majemuk. *Morfologi* dalam bahasa Inggris, hanya terdapat satu jenis imbuhan yaitu akhiran, sedangkan *morfologi* dalam bahasa Indonesia terdapat beberapa jenis imbuhan, yaitu : awalan, sisipan, akhiran dan gabungan awalan akhiran. Hal inilah yang membuat *stemming* di dalam bahasa Indonesia lebih rumit daripada *stemming* pada bahasa Inggris. Sehingga muncullah algoritma penghapusan imbuhan khusus bahasa Indonesia yang hingga saat ini kurang lebih telah terdapat 8 algoritma *stemming* bahasa Indonesia.

Diawali oleh penelitian Bobby Nazief dan Mirna Adriani pada tahun 1996, kemudian terdapat penelitian kembali dengan ditemukannya algoritma *stemming* Vega tahun 2001, selanjutnya Arifin dan Setiono pada tahun 2002 juga kembali melakukan penelitian terhadap *stemming* bahasa Indonesia, dan pada tahun 2005 Fadhila Tala melakukan penelitian penerapan algoritma Porter pada bahasa Indonesia, kemudian dilanjutkan penelitian kembali penelitian terhadap *stemming* oleh Jelita Asian pada tahun 2007 yang membandingkan 5 algoritma *stemming* sehingga muncul modifikasi algoritma Nazief Andriani yang biasa

disebut *CS*, setelah penelitian Jelita Asian tersebut *CS* dikembangkan kembali oleh Agus Zainal Arifin pada tahun 2008 sehingga memunculkan *ECS*, dan pada tahun 2011 *ECS* dikembangkan kembali oleh Andita Dwiyoga memunculkan algoritma *CC*. Penelitian terakhir mengenai *stemming* adalah pada tahun 2012 yaitu penelitian penggabungan algoritma *ECS* dengan Porter untuk bahasa Indonesia, sehingga algoritma terbaru dalam *stemming* adalah *ECSP*.

Dari 8 algoritma tersebut penghapusan imbuhan selalu melakukan penghapusan satu persatu jenis imbuhan, pemenggalan awal umumnya adalah pemenggalan *inflection suffixes* yang terdiri dari Partikel dan keterangan milik, kemudian melakukan penghapusan akhiran, dan selanjutnya adalah penghapusan awalan. Dengan proses yang detail tersebut tentunya akan mendapatkan hasil *stemming* yang sempurna, namun akan memakan waktu yang cukup lama.

Penelitian dalam Tugas Akhir ini ingin menanggulangi permasalahan tersebut yaitu mencoba merekonstruksi algoritma *stemming* pada bahasa Indonesia dengan teknik penghapusan imbuhan sekaligus tanpa sebuah proses yang panjang. Hal tersebut tentunya akan disesuaikan dengan aturan pola imbuhan bahasa Indonesia yang terdapat pada Kamus Besar Bahasa Indonesia yang sesuai pada Pedoman Umum Ejaan Yang Disempurnakan pada tahun 2009.

1.2. Perumusan masalah

Dari latar belakang yang telah dijelaskan pada penjelasan sebelumnya maka dapat dimunculkan rumusan masalah, yaitu bagaimana mengoptimalkan kecepatan dan ketepatan proses *stemming* pada kata dalam bahasa Indonesia.

1.3. Tujuan dan manfaat

Tujuan dari penelitian ini adalah untuk menghasilkan rekonstruksi *stemming* bahasa Indonesia berdasarkan pola imbuhan kata dalam *morfologi* bahasa Indonesia.

1.4. Batasan masalah

Batasan masalah di dalam penelitian ini adalah :

- Algoritma *Stemming* yang akan dibuat hanya dapat digunakan pada kata berbahasa Indonesia yang sesuai dengan Ejaan Yang Disempurnakan (EYD).
- Data percobaan menggunakan data yang digunakan pada penelitian sebelumnya yaitu pada penelitian *ECSP* oleh Muhammad Alif Teknik Informatika Universitas Trunojoyo Madura yaitu dokumen penelitian Fadhila Tala dan *Crawl* RISE.

1.5. Metodologi penelitian

Metodologi penelitian yang akan dipakai pada penelitian ini nantinya adalah studi literature, studi lapangan, analisa dan perancangan aplikasi, implementasi dan pengujian aplikasi, penarikan kesimpulan. Berikut alur beserta penjelasan dari masing-masing point di atas :

- Studi Literatur
Pada tahap ini dilakukan pembelajaran dari paper, buku, atau artikel yang terkait mengenai *stemming*, *morfologi* bahasa Indonesia, *regular expresion* dan *transition graph*.
- Studi Lapangan
Pada tahap ini di lakukan pengumpulan data untuk pengujian sistem, data tersebut adalah dari data-data pada penelitian sebelumnya agar dapat dijadikan acuan dan perbandingan keakuratan sistem.
- Analisa dan Perancangan Aplikasi
Menganalisa kebutuhan apa saja yang nantinya dibutuhkan untuk membuat aplikasi dalam penelitian ini seperti morfologi bahasa Indonesia untuk menentukan pola imbuhan.
- Pembuatan dan Pengujian Sistem
Mengimplementasikan rancangan sistem yang sudah di buat dalam bentuk sekaligus melakukan pengujian.
- Penarikan Kesimpulan
Merupakan langkah akhir yang nantinya berisi kesimpulan dari penelitian yang dilakukan dan menulisnya dalam laporan.

1.6. Sistematika penulisan

Laporan tugas akhir ini dibagi menjadi lima bagian pembahasan yaitu :

BAB I PENDAHULUAN

Pada bagian ini dipaparkan mengenai hal-hal yang melatarbelakangi penelitian, rumusan dan batasan permasalahan dalam penelitian, serta tujuan dan relevansi atau manfaat penelitian terhadap perkembangan solusi dari permasalahan, serta sistematika penulisan dalam laporan tugas akhir ini.

BAB II TINJAUAN PUSTAKA

Dalam bab ini diuraikan teori-teori penunjang dalam penulisan laporan tugas akhir ini. Pustaka yang digunakan baik dari buku, diktat, maupun materi dari internet.

BAB III PERANCANGAN SISTEM

Dalam bagian ini diuraikan secara rinci rancangan system dari aplikasi tugas akhir ini.

BAB IV IMPLEMENTASI SISTEM

Pada bagian ini diuraikan secara rinci implementasi program serta hasil analisa yang diperoleh.

BAB V KESIMPULAN

Pada bab ini dijelaskan hasil dari implementasi, yang termaktub di dalam kesimpulan. Dan juga saran untuk mengembangkan hasil analisis penelitian tugas akhir ini agar dapat dikembangkan ke penelitian berikutnya.