

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. Bertambahnya jumlah dokumen teks yang dapat diakses di internet diikuti dengan meningkatnya kebutuhan pengguna akan perangkat pencarian informasi yang efektif dan efisien, menyebabkan diperlukannya pengembangan dari teknologi IR. Efektif berarti user mendapatkan dokumen yang relevan dengan query yang diinputkan. Efisien berarti waktu pencarian yang sesingkat-singkatnya [1]. Salah satu penerapan dari *Information Retrieval (IR)* adalah *Search Engine*. Dengan adanya *Search Engine*, dokumen yang dicari dapat ditemukan sesuai dengan kebutuhan.

Dalam perkembangannya, kurang lebih ada sekitar 35 imbuhan resmi yang disebutkan dalam Kamus Besar Bahasa Indonesia. Imbuhan-imbuhan ini terdiri dari prefiks (awalan), sufiks (akhiran), konfiks, maupun infiks (sisipan). Satu hal yang unik dari Bahasa Indonesia adalah kecenderungan pemakaian imbuhan secara bebas dan imbuhan-imbuhan tersebut dapat dikombinasikan satu dengan lainnya [2].

Seperti halnya bahasa-bahasa lainnya, Bahasa Indonesia juga memiliki struktur sehingga bahasa tersebut dapat dipahami oleh banyak orang. Dalam Bahasa Indonesia, terdapat huruf-huruf, kata-kata, dan kalimat-kalimat dimana kumpulan huruf-huruf membentuk kata, kumpulan kata membentuk kalimat kumpulan kalimat membentuk paragraf, dan seterusnya. Pada penerapannya dalam tulisan, banyak orang yang masih salah saat menuliskan kata-kata dalam bahasa Indonesia sesuai dengan kata dasarnya.

Dalam *Information Retrieval (IR)*, *stemming* merupakan bagian dari *preprocessing* yang dimana fungsinya agar menemukan kata dasar dari sebuah kata dengan menghilangkan semua imbuhan. Untuk penentuan pemisahan dari imbuhan yang terdapat pada suatu kata maka pada *stemming* ini memiliki beberapa algoritma yang sesuai. Salah satu metode yang digunakan pada *stemming* Bahasa Indonesia adalah *ECS*

(*Enhanced Confix Stripping*) dan *Porter Stemmer*. *ECS* adalah stemming yang sudah teruji keakuratannya tetapi prosesnya lambat dan *Porter Stemmer* merupakan metode stemming paling cepat dalam pemrosesan data namun hasilnya tidak seakurat *ECS*.

RISE (Research Indonesian Search Engine) adalah project *search engine* terbuka yang sedang dikembangkan. Saat ini, *RISE* sudah menyelesaikan *web crawler* (penghimpun dokumen) dan masih banyak tahapan yang harus dilakukan seperti *preprocessing*. Tahapan *preprocessing* akan dikembangkan pada penelitian ini dengan fokus pada *stemming*. Penelitian ini mengusulkan digabungkannya metode *stemming ECS (Enhanced Confix Stripping)* dan *Porter Stemmer*.

1.2 Perumusan Masalah

Dari latar belakang yang sudah dijelaskan sebelumnya dimunculkan rumusan masalah yaitu

- a. Bagaimana cara menggabungkan metode stemming *ECS* dan *Porter Stemmer*?
- b. Bagaimana menghasilkan stemming yang efektif dan efisien hasil penggabungan metode *ECS* dan *Porter Stemmer*?

1.3 Tujuan

Tugas akhir ini akan membangun suatu *Stemmer* khusus konten Bahasa Indonesia sehingga *RISE* (pada tahapan *preprocessing*) mampu:

- a. Menggabungkan metode *ECS* dan *Porter Stemmer*.
- b. Menghasilkan stemming yang efektif dan efisien hasil penggabungan metode *ECS* dan *Porter Stemmer*.

1.4 Batasan Masalah

Pembatasan ruang lingkup penelitian, yaitu:

- a. Penelitian ini tidak menangani pembuatan *web crawler*, karena telah disediakan *web crawler* untuk konten Bahasa Indonesia pada penelitian sebelumnya.
- b. Hanya menggunakan kombinasi metode *ECS* dan *Porter Stemmer*.

- c. Pengujian kombinasi *Stemmer* akan menggunakan data penelitian Fadillah Z Tala dan data RISE serta uji coba kinerjanya dilakukan pada sebuah aplikasi sederhana, untuk penerapan serta penelitian lebih lanjut akan diterapkan pada antar muka pengguna RISE.

1.5 Metodologi Penelitian

Pembuatan tugas akhir ini terbagi menjadi beberapa tahap pengerjaan yang tertera sebagai berikut :

1. Studi Literatur dan Identifikasi Permasalahan
Tahap ini dilakukan dengan mempelajari konsep-konsep yang berkaitan dengan tugas akhir ini untuk memahami metode analisa kebutuhan, perancangan sistem dan pendukung lainnya melalui literature-literatur seperti buku, paper, sumber ilmiah lain seperti *internet* ataupun artikel dokumen teks lainnya yang berhubungan .
2. Analisa dan Perancangan sistem
Dalam tahap ini merupakan tahapan pertama dari pengembangan sistem. Pada tahap ini mendefinisikan permasalahan yang ada pada penelitian dan menentukan tujuan perancangan sebagai acuan untuk menentukan langkah-langkah selanjutnya. Perancangan sistem dan deskripsi sistem berdasarkan data hasil observasi menggunakan *flowchart*, diagram dan lain-lain.
3. Pembuatan Program
Tahap ini merupakan tahap pembuatan dan pengembangan *program* sesuai dengan perancangan sistem yang ditetapkan pada tahap sebelumnya.
4. Uji Coba dan Evaluasi
Melakukan ujicoba terhadap program yang sudah dibuat dengan skenario awal atau tujuan yang sudah ditentukan dan kemudian hasil uji coba tersebut di evaluasi prototipe perangkat lunak untuk mencari masalah yang mungkin timbul, mengevaluasi jalannya program, dan mengadakan perbaikan jika ada kekurangan dari program tersebut. Dilakukan

perbandingan *ECS*, *Porter Stemmer* dan *Stemmer* yang dikembangkan oleh penulis.

5. Dokumentasi

Pada Tahap ini dilakukan dokumentasi mengenai apa yang diperoleh dan dikerjakan mulai dari kegiatan studi literatur dan identifikasi permasalahan hingga menghasilkan program baru.

1.6 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Pada bab ini terdiri dari : latar belakang, perumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan.

BAB II : STUDI PUSTAKA

Pada bab ini dijelaskan mengenai hal-hal yang melandasi atau yang akan digunakan sebagai pedoman untuk menyelesaikan permasalahan yang terapat dalam tugas akhir ini.

BAB III : PERANCANGAN SISTEM

Pada bab ini akan dijelaskan analisis permasalahan, solusi dalam penanganan masalah, analisis kebutuhan dan perancangan sistem.

BAB IV : IMPLEMENTASI SISTEM

Pada bab ini akan dijelaskan implementasi aplikasi *Stemmer* untuk RISE yang ditentukan berdasarkan rancangan yang telah dibuat, kemudian melakukan ujicoba dan evaluasi terhadap program yang sudah diimplementasikan.

BAB V : PENUTUP

Pada bab ini berisikan kesimpulan dari hasil ujicoba serta saran yang dapat digunakan untuk pengembangan selanjutnya.