

BAB I

PENDAHULUAN

Bagian ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

1.1 Latar Belakang Masalah

Stemmer merupakan salah satu langkah yang digunakan dalam beberapa jenis aplikasi pengolahan data teks, seperti: *text mining*, *information retrieval (IR)*, dan *natural language processing (NLP)*. Tugas utama dalam *stemmer* adalah mengembalikan kata ke kata dasarnya (*root* atau *stem*) [1]. *Stemmer* bahasa Arab bukanlah tugas yang mudah.

Secara morfologis, bahasa Arab adalah salah satu bahasa yang paling rumit dan kaya. Puluhan atau ratusan kata dapat dibentuk dengan menggunakan satu kata dasar, beberapa pola dan beberapa afiks. Bahasa Arab juga memiliki tingkat ambiguitas yang tinggi karena berbagai alasan, seperti penghilangan huruf vokal dan kesamaan huruf tetap dengan huruf induk atau akar. Analisis morfologis biasanya mempengaruhi tingkat analisis yang lebih tinggi seperti analisis sintaksis dan semantik [2].

Beberapa jurnal dan proyek penelitian diusulkan untuk mengembangkan *stemmer* bahasa Arab (misalnya Al-Shalabi dan Evens, 1998: Khoja dan Garside, 1999: Abu-Salem, dll. 1999). Ada banyak penelitian yang menyajikan contoh algoritma *stemmer* bahasa Arab dan efektivitasnya. Sebagian besar penelitian ini mengklaim akurasi yang melebihi 85%. Akan tetapi tidak mungkin untuk memverifikasi klaim ini karena kurangnya *source codes* dan kumpulan data yang digunakan dalam proses pengujian [3].

Stemmer yang efektif dan efisien dalam hal ini dapat meningkatkan berbagai permasalahan yang sebelumnya ditemukan dalam proses pencarian informasi dan pengolahan data teks. Dalam penelitian ini penulis mencoba melakukan perbaikan *stemmer* bahasa Arab dengan mengevaluasi serta membandingkan hasil dari kinerja *light stemmer* dan *heavy stemmer*. *Data set* yang digunakan berupa kumpulan kata yang diambil dari *Al-Qur'an* surat *Al-Fil* s/d *An-Nas*.

Hal ini yang menjadi latar belakang penulis untuk mengkaji beberapa literatur yang berkaitan dengan *stemmer* bahasa Arab. Oleh karenanya skripsi tugas akhir ini penulis beri judul **“Perbaikan Algoritma Stemmer Bahasa Arab yang Diterapkan pada Kosakata Al-Qur’an”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, skripsi tugas akhir ini memiliki rumusan masalah, antara lain:

1. Bagaimana algoritma *stemmer* bahasa Arab?
2. Bagaimana kinerja dan hasil dari algoritma yang digunakan untuk *stemmer* bahasa Arab?
3. Bagaimana perbaikan yang dapat dilakukan untuk meningkatkan hasil *stemmer* bahasa Arab?

1.3 Batasan Masalah

Batasan masalah pada skripsi tugas akhir ini antara lain:

1. *Stemmer* yang digunakan yaitu terdiri dari *light stemmer*: *ARLStem stemmer*, *Snowball stemmer*, dan *heavy stemmer*: *Khoja stemmer*, *ISRI stemmer* dan *Al-Kabi stemmer*.
2. *Data set* yang digunakan yaitu kumpulan kata dalam *Al-Qur’an* surat *Al- Fil* sampai dengan *An-Nas* yang bersumber dari *Quranic Arabic Corpus*.
3. Bahasa pemrograman yang digunakan yaitu *Python*.
4. Perbaikan yang dilakukan berfokus pada perbaikan algoritma *heavy stemmer*.

1.4 Tujuan Penelitian

Adapun tujuan dan manfaat dari skripsi tugas akhir ini antara lain:

1. Memahami bagaimana proses *stemmer* bahasa Arab.
2. Membandingkan proses dari masing-masing algoritma *stemmer* bahasa Arab (*light stemmer* dan *heavy stemmer*).
3. Mengevaluasi hasil kinerja dari masing-masing algoritma *stemmer* bahasa Arab (*light stemmer* dan *heavy stemmer*).

4. Mengetahui bagaimana perbandingan *stemmer* bahasa Arab (*light stemmer* dan *heavy stemmer*).
5. Melakukan perbaikan pada *stemmer* bahasa Arab yang digunakan.

1.5 Metode Penelitian

Metode yang penulis tempuh dalam menyelesaikan skripsi ini adalah menggunakan pendekatan teoritis atau studi literatur. Penulis mencari dan mengumpulkan data berupa informasi yang mendukung pengerjaan skripsi ini, yaitu mengenai *stemmer*, *arabic stemmer*, *text mining*, *information retrieval (IR)*, dan *natural language processing (NLP)*. Sumber-sumber tersebut dapat diperoleh dari *website*, jurnal, skripsi, buku, dan lainnya. Adapun dalam penerapan model terhadap studi kasus dilakukan dengan menggunakan data sekunder yang diperoleh dari sumber yang sudah ada.

1.6 Sistematika Penulisan

Berdasarkan sistematika penulisan, skripsi ini terdiri atas lima bab ditambah dengan daftar pustaka, dimana setiap babnya memiliki beberapa subbab.

BAB I PENDAHULUAN

Bagian ini mengemukakan beberapa hal mengenai pendahuluan yang mendukung dalam penulisan skripsi ini. Berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bagian ini mengemukakan teori-teori yang melandasi pembahasan utama dari skripsi, yang meliputi pembahasan bahasa Arab, pengolahan teks, bahasa pemrograman dan package dan penelitian terkait.

BAB III ALGORITMA STEMMER BAHASA ARAB

Bagian ini berisi analisis algoritma *stemmer* bahasa Arab yang digunakan dalam skripsi, yang meliputi pembahasan mengenai algoritma *ARLStem stemmer*, *Snowball stemmer*, *Khoja stemmer*, *ISRI stemmer*, *Al-Kabi stemmer*, analisis afiks (prefiks dan sufiks), analisis pola kata, kekurangan dan kelebihan *stemmer*.

BAB IV PERBAIAKN DAN EVALUASI HASIL KINERJA STEMMER

BAHASA ARAB

Bagian ini berisi pembahasan utama dari skripsi, yang meliputi data, evaluasi hasil kinerja dari masing-masing algoritma *stemmer* bahasa Arab, dan perbaikan *stemmer* bahasa Arab.

BAB V PENUTUP

Bagian ini berisi simpulan sebagai hasil dari rumusan masalah yang telah dipaparkan serta berisi saran untuk penelitian selanjutnya sebagai pengembangan dari topik permasalahan yang bersangkutan.

DAFTAR PUSTAKA

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG