

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang penelitian yang dilakukan dan menjelaskan permasalahan serta solusi yang dilakukan.

1.1 Latar Belakang

Information Retrieval merupakan cara untuk mengambil dan menemukan informasi yang relevan dengan kebutuhan pengguna dari kumpulan informasi yang ada secara otomatis. *Information retrieval* berhubungan dengan pengambilan informasi dari dokumen yang tersedia berdasarkan *query* yang diinputkan sesuai keinginan user. Hasil dari *Information retrieval* yaitu dokumen – dokumen yang relevan dan kemudian diberikan kepada user[1]. Pemanfaatan teknologi informasi di era saat ini sudah banyak dampaknya di berbagai bidang, tidak terkecuali dalam melakukan pengelolaan dokumen. Akibatnya pertumbuhan jumlah dokumen digital tidak dapat dihindari, semakin banyak jumlah dokumen yang dimiliki maka berpengaruh terhadap kecepatan dan ketepatan dalam memperoleh informasi[2].

Stemming merupakan salah satu tahapan yang ada pada *information retrieval* yang dapat meningkatkan kinerja dari *information retrieval* dalam me-retrieve dokumen. Cara kerja *stemming* yaitu menghilangkan imbuhan dari sebuah kata menjadi bentuk kata dasarnya[3]. Pada *stemming* Bahasa Indonesia imbuhan yang dihapus berupa imbuhan awalan, akhiran dan sisipan[4]. Ada banyak algoritma *stemming* yang dikembangkan untuk *stemming* Bahasa Indonesia seperti algoritma Nazief & Adriani, algoritma Arifin & Setiono, dan algoritma *Porter Stemmer*[5]. Pada penelitian yang dilakukan oleh Dwi Wahyudi, dkk[6], membahas tentang perbandingan hasil dan waktu proses pada algoritma *stemming* Nazief & Adriani dan algoritma *Porter stemmer*. Didapatkan hasil algoritma *Porter stemmer* lebih cepat dibandingkan algoritma Nazief & Adriani dengan waktu proses yang dihasilkan yaitu 12.38 detik untuk *Porter stemmer* dan 22.16 detik untuk Nazief & Adriani. Algoritma Nazief & Adriani dikembangkan dengan menambahkan beberapa aturan morfologi yang kemudian diberi nama *Confix Stripping* (CS)[7]. CS masih memiliki kekurangan yang kemudian dikembangkan menjadi algoritma *Enhanced Confix Stripping* (ECS) *Stemmer*. ECS memiliki kinerja *stemming* yang paling baik dan kesalahan *stemming* yang lebih sedikit dibandingkan CS dalam melakukan *stemming* Bahasa Indonesia[8].

Al-Qur'an sejak diwahyukan sampai saat ini sudah diterjemahkan kedalam

bermacam-macam bahasa yang ada di dunia termasuk Bahasa Indonesia. Sudah banyak hasil karya Al-Qur'an yang diterjemahkan ke dalam Bahasa Indonesia. Al-Qur'an merupakan suatu sumber hukum yang menjadi pedoman hidup umat Islam dalam menjalani kehidupan bermasyarakat, beragama dan bernegara. Dalam kehidupan banyak permasalahan tentang agama yang sulit untuk dipecahkan. Untuk memudahkan dalam menyelesaikan permasalahan yang ada, dibutuhkan untuk memahami dan mengkaji Al-Qur'an. Ketika ingin membaca terjemahan Al-Qur'an agar tidak terjadi kesalahan ketika menafsirkan atau memahami makna yang terkandung didalamnya, dibutuhkan sebuah tafsir Al-Qur'an. Banyaknya ayat dan surat yang terdapat didalam Al-Qur'an yang membuat pencarian tafsir Al-Qur'an berdasarkan topik yang diinginkan membutuhkan waktu yang lama. Pencarian tafsir ayat Al-Qur'an digital yang ada di internet tidak cukup membantu dalam melakukan pencarian tafsir berdasarkan topik yang akan dicari[9][10].

Berdasarkan uraian latar belakang diatas. Pada penelitian sebelumnya, penggunaan *Porter Stemmer* dan Nazief & Adriani menunjukkan bahwa proses *stemming* lebih cepat menggunakan *Porter Stemmer*. Sedangkan pada penelitian yang lain, algoritma Nazief & Adriani telah dikembangkan menjadi *Confix Stripping* (CS) kemudian CS dikembangkan lagi menjadi *Enhanced Confix Stripping* (ECS) *Stemmer*, sehingga kinerjanya menjadi semakin baik. Oleh karena itu, dibuatlah sistem perbandingan nilai akurasi dan waktu proses pada algoritma *Porter Stemmer* dengan algoritma *Enhanced Confix Stripping Stemmer* untuk membuktikan apakah *Porter Stemmer* lebih cepat dibandingkan dengan *Enhanced Confix Stripping Stemmer* dalam melakukan pencarian teks pada tafsir Al-Qur'an berbahasa Indonesia.

1.2 Perumusan Masalah

1.2.1 Permasalahan (*Research Problem*)

Algoritma *Porter Stemmer* menunjukkan waktu prosesnya lebih cepat dibandingkan dengan algoritma Nazief & Adriani. Algoritma Nazief & Adriani dikembangkan menjadi CS. Namun, CS masih memiliki kekurangan yang kemudian dikembangkan lagi menjadi *ECS Stemmer*, sehingga metode *stemming* ECS menjadi semakin baik. Maka dari itu dilakukan penelitian untuk membandingkan apakah algoritma *Porter Stemmer* lebih cepat dibandingkan *ECS Stemmer*.

1.2.2 Metode Usulan (*Proposed Method*)

Metode yang diusulkan yaitu algoritma *Porter Stemmer* dan algoritma *ECS Stemmer* untuk melakukan pencarian teks pada tafsir Al-Qur'an berbahasa Indonesia.

1.2.3 Pertanyaan Penelitian

1. Berapa perbandingan akurasi pencarian teks menggunakan *Porter Stemmer* dan menggunakan *ECS Stemmer* pada *Cosine Similarity*?
2. Berapa perbandingan waktu proses pencarian teks menggunakan *Porter Stemmer* dan menggunakan *ECS Stemmer* pada *Cosine Similarity*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

1. Mengetahui perbandingan akurasi pencarian teks menggunakan *Porter Stemmer* dan menggunakan *ECS Stemmer* pada *Cosine Similarity*.
2. Mengetahui perbandingan waktu proses pencarian teks menggunakan *Porter Stemmer* dan menggunakan *ECS Stemmer* pada *Cosine Similarity*.

1.3.2 Manfaat Penelitian

Hasil dari penelitian ini dapat membandingkan dua algoritma *stemming* yaitu algoritma *Porter Stemmer* dan *ECS Stemmer* pada saat melakukan pencarian teks tafsir Al-Qur'an berbahasa Indonesia.

1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah :

1. Data yang digunakan dalam sistem adalah data Tafsir Al-Qur'an As-Sa'di berbahasa Indonesia menggunakan juz 29 yang terdapat 11 surat memiliki 431 ayat dan tafsir [11].
2. Uji coba sistem dilakukan dengan menginputkan 15 *query* yang berbeda.
3. Tidak ada batasan untuk jumlah kata yang di *stemming*.
4. Sistem yang dibuat berbasis web.

1.5 Sistematika Penulisan

Adapun sistematika penulisan yang digunakan dalam skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang, perumusan masalah, tujuan dan manfaat, batasan masalah, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Pada bab ini menjelaskan tentang landasan teori dan penelitian-penelitian sebelumnya. Landasan teori menjelaskan tentang teori yang menjadi dasar yang dapat membantu dalam melakukan penelitian dan penyusunan laporan penelitian. Teori tersebut secara garis besar menjelaskan tentang *information retrieval*, *pre-processing*, pembobotan TF-IDF, *cosine similarity*, terjemahan Al-Qur'an As-Sa'di dan *confusion matrix*.

BAB III METODE USULAN

Pada bab ini menjelaskan tentang algoritma *porter stemmer* dan *ECS stemmer*, arsitektur sistem, tahapan penelitian dan skenario pengujian sistem.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan tentang implementasi pembuatan aplikasi dan hasil dari pengujian sistem berdasarkan rencana yang telah di buat sebelumnya.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari penelitian yang menjawab pertanyaan penelitian serta berisi saran mengenai penelitian yang sudah dilakukan.