

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 (IR) merupakan pengambilan informasi yang tersimpan didalam dokumen yang berisikan konteks dari dokumen tersebut. Dengan berkembangnya teknologi sistem pencarian mengalami peningkatan dalam mendapatkan hasil yang sesuai dengan *query* (kata kunci) [1]. *IR* dapat dijumpai berupa aplikasi *mobile* dan *desktop*, namun dari kedua sistem itu lebih banyak ditemui dalam bentuk *desktop* [2]. Dalam *IR* umumnya memiliki dua tujuan yaitu *retrieval* dan *browsing*. *Retrieval* memiliki kegunaan sebagai pencarian informasi khusus dan fokus dalam beberapa permasalahan tergantung pengguna saat menginputkan *query*, sedangkan *browsing* memiliki kegunaan menjelajah informasi secara luas [3].

IR memiliki beberapa model pencocokan antara *term* (kata) dan *query* (kata kunci) yang dicari dalam sebuah dokumen, yaitu [4] : model *boolean*, model *vektor*, model *probabilistik*. Model vektor adalah suatu model untuk mewakili dokumen dan *query* sebagai vektor *similarity* (ruang vektor). Metode ini menggunakan cara kerja mengukur kemiripan antar dokumen yang akan dikonversi kedalam suatu model vektor untuk diukur dengan *query*. Beberapa metode dari ruang vektor *similarity* yang biasa digunakan, yaitu : *cosine similarity* dan *dice similarity*. *Cosine similarity* merupakan metode untuk menghitung tingkat kemiripan antar *query* dan dokumen, untuk mengukur kemiripannya menggunakan fungsi nilai *cosinus* dengan perpaduan rumus *cross product* dan *dot product* dari sudut antara dua vektor [5], sedangkan *dice similarity* mengukur kesamaan antar dokumen yang telah di *retrieve* menggunakan *query* sebagai acuan menghitung nilai kesamaan antar dokumennya, *dice similarity* merupakan panjang normalisasi *inner product* dari dua buah vektor [6].

Stemming merupakan suatu proses yang terdapat dalam *IR* yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Penggunaan algoritma *stemming* Nazief dan Adriani ini digunakan untuk teks berbahasa Indonesia yang memiliki morfologi (pembentukan kata) yang sangat luas daripada algoritma *stemming* lainnya adapun hasil

yang didapat kata yang didapat lebih baik dari algoritma porter untuk mendukung sistem temu informasi (Dwi Whayudi 2017). Algoritma ini hanya bisa digunakan untuk dokumen berbahasa Indonesia karena terdapat pemenggalan kata imbuhan berupa *suffix* dan *prefix* sedangkan untuk pemrosesan *stemming* pada dokumen berbahasa Inggris hanya menggunakan pemenggalan kata *suffix* [7].

Proses *stemming* Nazief dan Adriani menggunakan data berupa teks yang berbahasa Indonesia, salah satunya yang dapat digunakan yaitu Al-qur'an. Al-quran yang telah diterjemahkan dalam 35 bahasa yang ada di dunia salah satunya bahasa Indonesia [8]. Al-qur'an dalam bahasa Arab kepada Rasulullah. Al-Zarqani menjelaskan bahwa Al-qur'an maha segala-Nya, hal tersebut telah terbukti melalui tanda-tanda kebesaran Allah SWT. Dengan berkembangnya teknologi banyak aplikasi Al-qur'an dan tafsir dalam berbagai bahasa yang bertujuan untuk memudahkan publik untuk memahami dari setiap ayat yang ada. Akan tetapi, masih terdapat kendala dalam melakukan pencarian dengan topik yang diinginkan misalnya hasil pencarian yang muncul kurang sesuai yang diinginkan.

Oleh karena itu, untuk memudahkan dalam pencarian suatu topik Al-qur'an terjemahan yang sesuai dengan topik maka dibuat sebuah sistem pencarian teks Al-qur'an terjemahan berbahasa Indonesia dengan metode *dice similarity* dan *stemming* nazief dan adriani.

1.2 Perumusan Masalah

1.2.1 Permasalahan (*Research Problem*)

Hasil pencarian topik terjemahan Al-qur'an yang tidak sesuai dengan kata kunci yang diinginkan yang membuat publik tetap kesulitan dalam mencari suatu topik dalam sebuah dokumen.

1.2.2 Metode Usulan (*Proposed Method*)

Solusi yang diusulkan adalah membuat sebuah sistem dengan tahapan *pre-processing*, *stemming*, pembobotan *TF-IDF* dan perhitungan *dice similarity* yang diharapkan dapat menghasilkan hasil yang sesuai. Melalui tahap *pre-processing*, hasil dari *pre-processing* yang akan di *stemming* menggunakan algoritma Nazief dan Adriani dilanjutkan mengidentifikasi dan pembobotan dengan menggunakan algoritma *TF-IDF*, tahap terakhir yaitu menghitung kemiripan antar dokumen menggunakan algoritma *dice similarity*.

1.2.3 Pertanyaan Penelitian

1. Berapa nilai akurat penggunaan metode *dice similarity* dalam proses pencarian teks tafsir Al-qur'an ?
2. Bagaimana pengaruh penggunaan algoritma *stemming* Nazief dan Adriani dengan tanpa penggunaan algoritma *stemming* Nazief dan Adriani dalam menghasilkan output yang lebih relevan pada proses pencarian teks tafsir Al-qur'an ?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

1. Mengetahui nilai akurasi penggunaan metode *dice similarity* dalam proses pencarian data pada teks tafsir Al-qur'an.
2. Mengetahui pengaruh penggunaan algoritma *stemming* Nazief dan Adriani dengan tanpa penggunaan algoritma *stemming* Nazief dan Adriani dalam menghasilkan output yang lebih relevan pada proses pencarian teks tafsir Al-qur'an

1.3.2 Manfaat Penelitian

Hasil dari penelitian ini dapat dijadikan rujukan untuk membangun sistem pencarian teks terjemahan Al-qur'an berbahasa Indonesia untuk memudahkan dalam mencari topik yang diinginkan.

1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah :

1. Data yang digunakan untuk penelitian ialah data Al-qur'an juz 29.
2. Tafsir yang dijadikan acuan ialah tafsir As-Sa'di
3. Sistem ini berbasis pencarian teks.
4. Sistem ini berbasis web.
5. Input *query* menggunakan teks bahasa Indonesia.
6. Menggunakan Bahasa pemrograman python.
7. Menggunakan Framework Flask

1.5 Sistematika Penelitian

Adapun sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini dibahas tentang latar belakang mengenai ruang lingkup masalah yang dibahas, solusi yang diberikan untuk dapat menyelesaikan permasalahan, tujuan dan

manfaat dengan diterapkannya solusi yang diusulkan, batasan solusi yang diberikan, serta sistematika penulisan.

BAB II KAJIAN PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut diantaranya adalah penelitian- penelitian terkait identifikasi teks tafsir Al-qur'an, *TF-IDF*, *stemming* Nazief dan Adriani, dan *dice similarity*.

BAB III METODE USULAN

Pada bab ini dijelaskan mengenai rancangan dari sistem yang akan dibuat setelah melakukan analisa masalah, mulai dari rancangan model sistem, serta alur dari sistem yang akan dikembangkan.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini akan menjelaskan implementasi pembuatan aplikasi berdasarkan hasil rancangan yang telah di buat sebelumnya.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang disesuaikan dengan pertanyaan penelitian dan saran mengenai metode yang digunakan.