

BAB I PENDAHULUAN

1.1 Latar Belakang

Dunia tempat kita hidup ini banyak sekali menyimpan informasi-informasi yang bisa kita gunakan untuk kebutuhan penelitian. Pada dasarnya kita sendiri tidak bisa lepas dari yang namanya proses pencarian informasi. Pada kondisi apapun, pasti akan selalu ada model yang dibutuhkan untuk melakukan proses pencarian tersebut. Dalam dunia komputer, proses pencarian menjadi suatu proses yang sangat vital, dimana semua kegiatan *information technology* selalu berkaitan dengan hal tersebut. Di negara kita tercinta ini, teknologi informasi sudah berkembang dengan pesat. Masyarakat sudah banyak menggunakan teknologi informasi untuk berbagai keperluan, hal ini mengakibatkan meningkatnya efisiensi dan efektifitas yang signifikan dalam penyampaian informasi.

Hal inilah yang membuat *Information Retrieval* menjadi salah satu teknologi yang sangat dibutuhkan untuk proses pencarian informasi yang cepat dan akurat. Ada banyak sekali sistem informasi yang mampu menyimpan beragam informasi dalam bentuk teks. Karena dokumen yang tersimpan dalam sebuah sistem jumlahnya bisa mencapai ribuan, ratusan bahkan jutaan, maka proses pencarian akan menjadi aktifitas yang menguras banyak waktu. Mencari dokumen satu per satu bukanlah solusi yang baik mengingat banyaknya dokumen yang harus di periksa. Oleh karena itu, dibutuhkan sebuah aplikasi berupa alat yang mampu menemukan informasi yang sesuai dengan kebutuhan pengguna dengan waktu secepat mungkin. Hal ini yang melatar belakangi penciptaan teknologi *information retrieval*, dimana *information retrieval* menawarkan proses pencarian dokumen dengan cepat dan tepat.

Information Retrieval disini memiliki beberapa model yang dapat diaplikasikan pada data dokumen yang banyak tersebut, model yang pertama adalah *Boolean Model*. Model *Boolean* mengubah kalimat dalam dokumen sebagai suatu himpunan kata kunci (*set of keywords*). Lalu untuk *query* diekspresikan dalam bentuk *Boolean*. *Query* dihubungkan dengan operasi logika *boolean* seperti *AND*,

OR dan *NOT* yang mana terdapat pula tanda kurung untuk menentukan *query* mana yang akan diproses terlebih dahulu. Diharapkan dari pencarian ini diperoleh beberapa dokumen yang relevan dengan *query* tersebut [9].

Pada penelitian ini penulis tertarik untuk membuat sebuah program *Information Retrieval* dengan menggunakan metode *Boolean Model* dan *Vector Space Model* yang diaplikasikan menggunakan data surah Al-Qur'an dan hadis. Harapannya program ini kelak bisa digunakan oleh *user* secara *offline* sehingga *user* tetap bisa mendapatkan manfaatnya sekalipun sedang tidak terhubung dengan internet. Pada penelitian ini terdapat tiga skenario percobaan yang mana pada skenario pertama *dataset* terdiri dari 6236 ayat dan 30845 hadis. Untuk skenario kedua *dataset* terdiri dari 114 surah dan 30845 hadis sedangkan pada *dataset* ketiga terdiri dari 802 tema. Pada penelitian ini, akan dibahas bobot yang dihasilkan oleh masing-masing metode serta durasi *runtime* yang dibutuhkan oleh metode untuk melakukan satu kali proses pencarian.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah penulis sampaikan sebelumnya, penulis merumuskan masalah dalam penulisan skripsi ini, yaitu:

1. Menganalisis perbedaan total bobot yang dihasilkan antara metode *Boolean Model* dan *Vector Space Model* dengan tiga *dataset* berbeda.
2. Mengukur *Runtime* yang dibutuhkan masing-masing metode untuk mengeksekusi tiga skenario berbeda.
3. Menganalisis pengaruh ukuran data dalam perhitungan *TF-IDF*.

1.3 Batasan Masalah

Agar tidak menyimpang dari pembahasan, maka ada beberapa batasan masalah dalam skripsi ini, sebagai berikut:

1. Terdapat tiga *dataset* yang digunakan, yakni *dataset* pertama (6236 ayat + 30845 hadis), *dataset* kedua (114 surah + 30845 hadis) serta *dataset* ketiga yaitu berisi 802 tema surah Al-Qur'an.
2. Variabel yang diperhatikan adalah variabel bobot dan waktu.
3. Program diimplementasikan pada bahasa pemrograman *python*.

4. Metode *Information Retrieval* yang digunakan adalah *Boolean Model* dan *Vector Space Model*.

1.4 Tujuan Penelitian

Adapun tujuan dan manfaat dari penelitian skripsi ini diantaranya:

1. Membuat sebuah program untuk melakukan pencarian informasi pada data Al-Qur'an dan hadis.
2. Mengetahui hasil percobaan dari metode *Boolean Model* dan *Vector Space Model* jika diberikan tiga *dataset* yang berbeda.
3. Mengetahui perbandingan sistem jika menggunakan *dataset* surah, ayat dan tema.

1.5 Metode Penelitian

Metodologi dalam penelitian skripsi ini diantaranya:

1. Studi Literatur
Pada tahap ini dilakukan pengumpulan fakta-fakta dan teori penunjang yang berhubungan dengan metode *Information Retrieval* dengan menggunakan metode *Boolean Model* dan *Vector Space Model*.
2. Analisis
Pada tahap ini penulis menganalisis dan melakukan penelitian terhadap metode *Boolean Model* dan metode *Vector Space Model*.
3. Simulasi
Pada tahap ini dilakukan simulasi *Boolean Model* dan *Vector Space Model* menggunakan *Google Colab* serta *Jupyter Notebook* agar didapatkan data hasil *output* untuk dilakukan perbandingan.

1.6 Sistematika Penulisan

Berdasarkan sistematika penulisannya, tugas akhir ini terdiri atas lima bab serta daftar pustaka, dimana dalam setiap bab terdapat beberapa subbab.

BAB I : PENDAHULUAN

Pada bab ini memuat latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan metode penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini berisi uraian materi tentang hal-hal yang melandasi pembahasan masalah dan teori-teori yang digunakan sebagai pedoman untuk menyelesaikan permasalahan. Meliputi *Information Retrieval*, *Boolean Model*, *Vector Space Model*, *TF-IDF*, serta beberapa *tools* yang digunakan.

BAB III : *INFORMATION RETRIEVAL* MENGGUNAKAN *BOOLEAN MODEL* DAN *VECTOR SPACE MODEL*

Pada bagian ini akan dijelaskan metodologi penelitian yang dilakukan, dimulai dari pengambilan data, *text preprocessing*, proses *TF-IDF*, lalu data yang telah melalui *TF-IDF* dikirim ke dalam metode *Boolean Model* dan *Vector Space Model* untuk didapatkan bobot tiap dokumen yang relevan dengan *query* yang selanjutnya akan diurutkan dari bobot tertinggi hingga terendah sebelum di-*retrieve* kepada *user*.

BAB IV : ANALISIS HASIL IMPLEMENTASI DAN PERCOBAAN MENGGUNAKAN *BOOLEAN MODEL* DAN *VECTOR SPACE MODEL* PADA DATA AL-QUR'AN DAN HADIS

Pada bab ini akan dijelaskan tentang analisis hasil implementasi dan percobaan *information retrieval*

menggunakan metode *Boolean Model* dan *Vector Space Model* pada data surah Al-Qur'an dan hadis.

BAB V : PENUTUP

Pada bab ini dibuatkan kesimpulan hasil dari percobaan yang dilakukan lalu terdapat beberapa saran untuk menambah performa dari program yang dibuat.

