

BAB I

PENDAHULUAN

1.1 Latar Belakang

Untuk mendukung pelaksanaan pendidikan di bangku perkuliahan, telah diterbitkan Buku Pedoman Pendidikan Program Teknologi Informasi dan Ilmu Komputer untuk civitas akademika, pegawai dan masyarakat luas, khususnya bagi mahasiswa baru Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya, yang bertujuan untuk memberikan gambaran tentang tata cara proses pendidikan. Adapun dalam pemanfaatannya masih memiliki beberapa keterbatasan, oleh karena itu di dalam penelitian skripsi ini penulis berusaha untuk mengimplementasikan suatu *Question Answering System* terhadap kueri pertanyaan yang diberikan berdasarkan pada Buku Pedoman Pendidikan tersebut.

Question answering system (QA system) adalah sistem yang mengizinkan user menyatakan kebutuhan informasinya dalam bentuk *natural language question* (pertanyaan dalam bahasa alami), dan mengembalikan kutipan teks singkat atau bahkan frase sebagai jawaban. Ketersediaan sumber informasi yang luas dan bervariasi, serta adanya perkembangan yang pesat dari teknik *Natural Language Processing (NLP)*, *Information Extraction (IE)*, dan *Information Retrieval (IR)* sangat mempengaruhi perkembangan dari *QA system*, yang mana dulunya hanya dapat menjawab pertanyaan-pertanyaan yang terbatas pada suatu bidang (domain) tertentu dengan berdasarkan pada sumber informasi yang terstruktur seperti database, hingga kini dapat menjawab berbagai jenis pertanyaan dengan bersumber pada informasi dari sebuah koleksi teks yang tidak terstruktur [GUN-06].

Question answering system (QA system) diterapkan dengan memasukkan kueri berupa pertanyaan ke dalam sistem, maka sistem akan mengembalikan suatu kalimat sebagai jawabannya. Proses penemukembalian jawaban dimulai dengan memecah (*parsing*) suatu dokumen menjadi kalimat-kalimat. Kalimat-kalimat tersebut dipecah dan di-stem menjadi *token-token*. Begitu pula dengan kalimat

pertanyaan pada kueri dipecah dan di-stem menjadi *token-token*. *Token-token* tersebut diproses dalam *rule* sesuai dengan tipe pertanyaannya, yaitu : APA, SIAPA, KAPAN, DIMANA, dan BERAPA. Di dalam *rule* akan diberikan nilai untuk masing-masing kalimat dokumen. Kalimat yang memiliki nilai tertinggi akan dikembalikan sebagai kalimat jawaban. Semakin besar hasil persentase jumlah kalimat relevan terhadap jumlah kalimat yang ditemukembalikan, maka kinerja sistem akan semakin baik.

Penelitian - penelitian tentang *QA System* dalam perkembangannya sudah diimplementasikan oleh Ballesteros dan Xiaoyan Li (2007) berupa *CL-QAS* yang digunakan untuk bahasa Inggris dan Mandarin. Dalam penelitian tersebut digunakan pembobotan *heuristic* dan *syntactic* untuk mengidentifikasi kandidat kalimat yang relevan. Cidhy (2009) mengembangkan *QA System* dengan menggunakan pembobotan *heuristic* ke dalam dokumen berbahasa Indonesia. Kemudian Fandi (2011) melakukan penelitian pada *QA System* dengan menggunakan metode *N-Gram* untuk pembobotan *passagenya*. Sedangkan Resnu (2012) telah dilakukan penelitian Implementasi *Pattern Based Approach* Pada *Question Answering System* untuk menjawab kueri pertanyaan pada Buku Pedoman Pendidikan Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya. Pada penelitian tersebut, kandidat jawaban melalui proses *filtering* dan *scoring* untuk membuang kandidat yang tidak sesuai. Pola jawaban digunakan untuk mengekstrak potongan kalimat yang menjadi jawaban pada kandidat jawaban.

Pada penelitian skripsi ini, penulis menerapkan Implementasi *Question Answering System* pada dokumen teks bacaan berbahasa Indonesia. Pada tahap *preprocessing* untuk membersihkan dan mengurangi data dari *noise*, tahapan yang dilalui yaitu *case folding*, *tokenizing*, *filtering* dan *stemming*. Pada proses *stemmingnya* menggunakan algoritma *Stemming Arifin-Setiono*. Pencarian jawaban dilakukan dengan pemberian *scoring* menggunakan pembobotan heuristik *Rabin-Karp* dan pembobotan *Rule-Based* yang keduanya nanti akan dibandingkan hasilnya. Kemudian dipilih jawaban berupa kalimat pada dokumen yang memiliki nilai skor tertinggi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan, dapat dirumuskan permasalahan dalam penelitian ini adalah sebagai berikut :

1. Bagaimana implementasi penggunaan Algoritma *Rabin-Karp* dalam pengembangan *Question Answering System*.
2. Bagaimana pengaruh dari penggunaan *stemming Arifin-Setiono* pada implementasi *Question Answering System* dengan algoritma *Rabin-Karp* dan *Rule-Based*.
3. Bagaimana perbandingan hasil implementasi *Question Answering System* Untuk Dokumen Teks Bacaan Berbahasa Indonesia Menggunakan Algoritma *Rabin-Karp* dibandingkan dengan Menggunakan Algoritma *Rule-Based*.
4. Bagaimana perbandingan nilai heuristik dan waktu proses dalam melakukan pemilihan jawaban yang relevan pada Algoritma *Rabin-Karp* dengan menggunakan *kgram* dan *modulo* yang berbeda.

1.3 Batasan Masalah

Untuk memfokuskan penelitian yang akan dilakukan, permasalahan yang ada dibatasi sebagai berikut :

1. Dokumen yang digunakan untuk penelitian ini adalah dokumen teks yakni Buku Pedoman Pendidikan Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya Tahun Akademik 2012/2013.
2. Sistem tidak memperhatikan kesalahan ejaan pada dokumen dan mengabaikan adanya sinonim/persamaan kata.
3. Algoritma yang digunakan untuk proses *stemming* adalah Algoritma *Arifin-Setiono*.
4. Pada proses *stemming* mengabaikan imbuhan sisipan.
5. Algoritma yang digunakan untuk proses pencocokan string adalah Algoritma *Rabin-Karp*.
6. Metode yang digunakan untuk *scoring* jawaban adalah metode pembobotan heuristic dan metode *Rule-Based*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk merancang sistem aplikasi dan mengimplementasikan metode *Rabin-Karp* pada *Question Answering System* menggunakan kueri pertanyaan berbahasa Indonesia pada Buku Pedoman Pendidikan Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik untuk penulis dan pembacanya. Manfaat yang diharapkan adalah sebagai berikut :

1. Sebagai media untuk pengimplementasian ilmu pengetahuan teknologi khususnya di bidang *Information Retrieval*.
2. Mendapatkan pengetahuan dan wawasan tentang *preprocessing*, *stemming* *Arifin-Setiono*, pembobotan *heuristik*, algoritma *Rabin-Karp* dan *Rule-Based* yang diterapkan pada *Question Answering System*.
3. Membantu institusi pendidikan untuk mendukung pemanfaatan buku pedoman pendidikan agar lebih efisien.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I PENDAHULUAN

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Menjelaskan tentang dasar teori dan referensi dalam melakukan implementasi *Question Answering System* pada tahap - tahap *preprocessing*, metode pencocokan *string*, pembobotan nilai untuk menemukan jawaban yang relevan, dan beberapa teori dasar dalam sistem informasi *retrieval*.

BAB III METODOLOGI DAN PERANCANGAN SISTEM

Membahas metode yang digunakan dalam penelitian. Terdiri dari studi literature, perancangan sistem, implementasi algoritma, pengujian dan analisa hasil implementasi, penulisan laporan.

BAB IV IMPLEMENTASI

Membahas tentang implementasi metode yang digunakan dalam *Question Answering System* dan uji coba terhadap program yang telah dibuat

BAB V PENGUJIAN DAN ANALISIS

Membahas proses dan melakukan analisa terhadap hasil pengujian algoritma pada *Question Answering System*.

BAB VI PENUTUP

Berisi tentang kesimpulan yang didapat dari pembuatan skripsi ini dan saran-saran yang mungkin dapat berguna dalam penelitian lebih lanjut.

