

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan perkembangan informasi dengan melimpahnya data mengakibatkan muncul keadaan yang disebut *information overload* alias *infobesity* yaitu keadaan yang merujuk kepada kesulitan seseorang dalam membuat keputusan dalam permasalahan yang disebabkan oleh informasi yang terlalu banyak (Toffler 1970). Hal umum yang menyebabkan terjadinya *information overload* diantaranya adalah pesatnya peningkatan produksi informasi baru serta duplikasi data.

Dalam upaya untuk menanggulangi permasalahan *information overload* tersebut dilakukanlah proses yang dinamakan *information retrieval* (IR) yang merupakan sebuah usaha untuk memperoleh informasi yang tepat dari koleksi sumber informasi. Salah satu metode yang sedang berkembang dalam IR adalah *Question Answering* (QA). QA merupakan disiplin ilmu komputer dalam bidang *information retrieval* dan *Natural Language Processing* (NLP) yang berkaitan dengan pembangunan sistem yang secara otomatis menjawab pertanyaan yang diajukan oleh pengguna dalam bahasa alami. Tujuan utamanya adalah memberikan jawaban singkat melalui masukan bahasa alami yang didukung oleh dokumen yang mendasarinya (Cooper & Rigger 2000).

Sejalan dengan semakin berkembangnya penelitian dalam QA, semakin banyak pula berbagai metode yang dikembangkan serta diterapkan. Berdasarkan

bentuk jawaban yang dihasilkan dapat dibedakan menjadi beberapa macam yakni jawaban berupa frase, klausa, kalimat serta paragraf.

Berkaitan dengan pembentukan jawaban dengan jenis *sentence-level* yaitu QA yang mengembalikan kalimat sebagai jawabannya terdapat beberapa metode yang telah dikembangkan diantaranya penelitian yang dilakukan oleh Grisham dan Sundheim (Grisham & Sundheim 1996) yakni pengembangan model *Named Entity Recognition* yaitu dengan melakukan penyisipan tag pada kata atau frase sesuai dengan jenis entitasnya dalam sebuah kalimat. Seperti kalimat “Soekarno lahir di Surabaya” maka akan mendapatkan tag <nama>Soekarno</nama> lahir di <lokasi>Surabaya</lokasi>.

Berbeda dengan penelitian yang dilakukan sebelumnya beberapa penelitian menitikberatkan pada pembelajaran pola seperti yang dilakukan oleh Eduard Hovy dkk (Hovy et al 2002) yang membangun pola jawaban berdasarkan kebiasaan kemunculannya seperti “Kapan X lahir” jawaban yang biasa muncul adalah “X lahir pada tahun 1950” sehingga didapat pola “X lahir pada tahun <Jawaban>”.

Basis Aturan merupakan salah satu metode dalam QA. Metode Basis Aturan memberikan pembobotan terhadap kandidat jawaban berdasarkan tipe pertanyaan yang digunakan yang dikemukakan oleh Ellen Riloff dan Michael Thelen (Riloff & Thelen 2000). Beberapa tipe pertanyaan yang digunakan yakni *siapa, kapan, mana, mengapa* dan *apa*. Setiap tipe pertanyaan memiliki aturan dan nilai masing-masing. Kandidat jawaban yang memiliki bobot tertinggi maka akan dikembalikan sebagai jawaban.

Pada dasarnya metode berbasis aturan dibangun untuk pengujian terhadap kemampuan membaca (*reading comprehension*). Aturan yang dibangun sangat sederhana namun mampu memberikan petunjuk untuk memperoleh jawaban yang diharapkan sehingga pada kenyataannya metode berbasis aturan tersebut dapat diterapkan di berbagai domain. Salah satu kelemahan metode ini adalah memproses seluruh kalimat yang berada dalam dokumen sumber. Sehingga kalimat yang tidak memiliki informasi tentang jawaban yang diharapkan namun memenuhi aturan yang dibangun maka kalimat tersebut akan tetap dilakukan pembobotan.

Dalam penelitian ini penulis akan melakukan pengujian hasil penerapan QA dengan metode berbasis aturan dengan sumber informasi dari berbagai dokumen yang berisi berita, cerita atau penjelasan. Guna menanggulangi kelemahan pada metode basis aturan yang disebutkan sebelumnya maka penulis mengkolaborasikan teknik seleksi pada level kalimat yang menekankan pada perbandingan jumlah kata kunci pada pertanyaan dan kandidat jawaban yang dikemukakan oleh Zhiping Zheng (Zheng 2005).

Dengan digunakannya sistem berbasis QA diharapkan terdapat interaksi antara pengguna serta sistem. Untuk memperoleh informasi pengguna memberikan masukan pertanyaan melalui bahasa alami dan hasil informasi akan mengikuti perubahan konten atau dokumen yang mendasarinya secara otomatis. QA menghasilkan jawaban berupa kalimat, berbeda dengan search engine yang mengembalikan daftar dokumen yang harus disaring kembali oleh pengguna untuk menemukan jawaban.

1.2 Rumusan Masalah

Dalam penelitian ini berdasarkan latar belakang diatas rumusan masalah yang akan dibahas adalah sebagai berikut:

1. Apakah QA dengan metode Basis Aturan dapat memberikan penilaian sesuai aturan yang dibangun;
2. Apakah seleksi kalimat yang digunakan mampu menyeleksi kalimat berdasarkan perbandingan kata kunci;
3. Bagaimana tingkat keakuratan jawaban yang diberikan oleh sistem terhadap pertanyaan yang diajukan berdasarkan penilaian dari sisi manusia (*human-sense*).

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Memperoleh jawaban berdasarkan aturan yang dibangun;
2. Memperoleh hasil seleksi kalimat berdasarkan penyeleksi kalimat pada sistem QA yang dibangun;
3. Memperoleh tingkat akurasi jawaban berdasarkan pertanyaan yang diajukan berdasarkan penilaian dari sisi manusia (*human-sense*).

1.4 Batasan Masalah

Adapun batasan masalah yang dirancang dalam penelitian ini adalah sebagai berikut:

1. Sistem *Question Answering* yang dibangun menggunakan pendekatan Basis Aturan;
2. Seleksi kalimat yang digunakan menggunakan pendekatan perbandingan jumlah kata kunci pada pertanyaan dan kandidat jawaban;
3. Proses *Natural Language Processing* yang digunakan hanya sebagai Analisis Morfologi;
4. Masukan pertanyaan yang akan diproses harus didahului dengan kalimat tanya seperti *siapa, kapan, mana(dimana, kemana), mengapa* dan *apa*;
5. Bahasa pertanyaan yang digunakan adalah bahasa Indonesia yang baku dan tidak menggunakan singkatan;
6. Pertanyaan yang diproses bukan merupakan frasa tunggal atau terdiri dari satu buah kata;
7. Hasil yang diberikan berupa potongan kalimat.

1.5 Manfaat Penelitian

Beberapa manfaat yang diperoleh dari penelitian yang diperoleh adalah sebagai berikut:

1. Mengetahui cara kerja sistem *Question Answering* dalam penerapannya di bidang *Natural Language Processing*;
2. Mengetahui cara kerja Basis Aturan guna memberikan penilaian dalam QA;
3. Mengetahui cara kerja seleksi kalimat yang digunakan;

4. Mengetahui proses-proses yang terdapat dalam bidang *Natural Processing Language*;
5. Hasil penelitian ini dapat dikembangkan lebih lanjut menggunakan metode-metode lainnya.

1.6 Metodologi

Berikut metodologi yang digunakan dalam proses pengerjaan penelitian ini adalah sebagai berikut:

1.6.1 Tahap Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah dengan studi literatur yang diambil dari buku, jurnal, serta internet dengan materi seputar *Question Answering Sistem* yang dititik beratkan pada penggunaan Metode Basis Aturan serta analisa dasar dari *Natural Language Processing*.

1.6.2 Tahap Pembuatan Perangkat Lunak

Tahapan-tahapan dalam proses pembuatan perangkat lunak adalah sebagai berikut:

1. Analisis dan perancangan perangkat lunak

Merupakan tahapan penganalisaan hal-hal yang diperlukan dalam pelaksanaan proyek pembuatan perangkat lunak ini mulai dari bahasa pemrograman yang digunakan, struktur data hingga keluaran yang diharapkan.

2. Implementasi

Detail mengenai implementasi program dilakukan sesuai hasil analisis pada tahapan sebelumnya.

3. Pengujian

Tahapan pengujian hasil pengkodean agar sesuai dengan rancangan yang telah ditentukan.

4. Hasil akhir dan penarikan kesimpulan.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam karya ilmiah ini adalah sebagai berikut:

- **BAB 1 PENDAHULUAN**

Bab ini menguraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan serta manfaat penelitian dan sistematika penulisan.

- **BAB 2 TINJAUAN PUSTAKA**

Bab ini memaparkan beberapa hal mengenai teori-teori dasar yang mendukung dan berhubungan dengan penelitian ini.

- **BAB 3 METODE PENELITIAN**

Bab ini merupakan penjabaran dari metode pengembangan QA sistem sampai dengan implementasinya.

- **BAB 4 PEMBAHASAN DAN HASIL PENELITIAN**

Pada bagian ini akan dibahas secara mendalam hal-hal yang akan menjawab apa yang sudah dirumuskan dalam rumusan masalah.

- **BAB 5 KESIMPULAN DAN SARAN**

Kesimpulan merupakan jawaban atas pertanyaan-pertanyaan pada sub-bab rumusan masalah dan saran merupakan kumpulan saran dan rekomendasi dari penulisan untuk penelitian dan pengembangan sistem lebih lanjut.