

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi yang berkembang dengan cepat menyebabkan banyak perubahan pada kehidupan kita. Perubahan tersebut membuat kita memasuki masa yang disebut era *digital*. Pada era *digital* ini banyak dikembangkan berbagai macam media yang mampu menghantarkan kita pada informasi yang tidak terbatas. Informasi yang tidak terbatas tersebut dapat kita peroleh salah satunya adalah dari internet.

Banyaknya informasi yang ada membuat masyarakat sulit menemukan informasi yang tepat yang sesuai dengan apa yang diinginkan. Kebanyakan masyarakat belum bisa mencari *keyword* yang tepat untuk mencari informasi yang diinginkan. Salah satu cara untuk mendapat informasi yang diinginkan adalah dengan mengetahui aspek-aspek penting pada informasi yang ada. Untuk mengetahui inti dari informasi tersebut dibutuhkan pemahaman kita terhadap suatu bacaan/artikel yang disebut dengan *reading comprehension*. Seperti contohnya ketika kita membaca novel, kita membutuhkan pemahaman yang kuat untuk mengetahui isi dari cerita yang dipaparkan dalam novel tersebut. Tentunya untuk mengetahui aspek-aspek penting tersebut perlu diajukan pertanyaan-pertanyaan agar dapat membantu kita menentukan informasi yang akan dicari. Karena pencarian informasi dengan pertanyaan membuat hasil yang diperoleh semakin spesifik. Oleh karena itu dibutuhkan sistem yang dapat mengembalikan informasi-informasi yang kita cari dengan *query* berupa pertanyaan sebagai salah satu cara untuk mendapatkan informasi yang spesifik.

Question answering adalah salah satu topik dari *Information retrieval* yang ditujukan untuk memperoleh jawaban yang tepat dari *query* yang berbentuk pertanyaan. Sebuah *question answering system* untuk *reading comprehension* harus dirancang untuk menganalisis suatu bagian dari bahasa alami secara otomatis dan menghasilkan jawaban untuk setiap pertanyaan berdasarkan informasi dalam bagian teks tersebut [1]. Jadi, tujuan utama *Question Answering* (QA) yaitu menampilkan jawaban atas pertanyaan yang diberikan pengguna. Dengan memasukkan *query* berupa pertanyaan, akan diperoleh sebuah kalimat sebagai jawabannya. Misalnya dengan *query* “Siapa ketua Mahkamah Agung?”, kalimat yang mungkin akan dikembalikan sebagai jawaban adalah: “*Selain menjadi penghubung PBNU dengan Mahkamah Agung (MA), Mahfud MD saat menjadi Ketua Mahkamah Konstitusi (MK) juga diminta tolong untuk perkara lain.*” [2]. Tentunya sesuai dengan bacaan mana yang dipilih. Dengan bacaan yang berbeda, tentu kalimat yang dikembalikan sebagai jawaban pun akan berbeda. Sudah ada penelitian yang dilakukan terhadap sistem *question answering* khususnya untuk *reading comprehension*. Seperti penelitian yang dilakukan Riloff dan Thelen [3] dengan menggunakan bahasa Inggris, Xiaoyan Hao dan kawan-kawan [1] dengan menggunakan bahasa mandarin dan Abufardeh dan kawan-kawan [4] dengan menggunakan bahasa Arab. Ketiga penelitian tersebut menggunakan *rule-based* sebagai metode yang dilakukan untuk mengevaluasi

sistem *question answering* untuk *reading comprehension*. Menurut Riloff dan Thelen, metode *rule-based* dipilih karena metode *rule-based* menggunakan leksikal dan semantic [3] untuk mendeteksi informasi penting yang dibutuhkan untuk menjawab pertanyaan yang diajukan terutama dalam kasus *reading comprehension test*.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dirumuskan masalah sebagai berikut:

- a. Bagaimana implementasi *question answering system* untuk *reading comprehension test* dalam Bahasa Indonesia dengan metode *rule-based*?
- b. Bagaimana analisis pengaruh bahasa Indonesia pada implementasi *question answering system* untuk *reading comprehension test* dengan metode *rule-based*?

1.3. Batasan Masalah

Adapun batasan masalah yang dilakukan agar pengerjaan Tugas Akhir ini tidak terlalu luas adalah:

- a. Penelitian ini mengacu pada penelitian yang telah dibuat oleh Akour, Abufardeh, Magel dan Al-Radaideh tentang *Question Answering for Reading Comprehension in Arabic* [4] yang sebelumnya telah dikembangkan oleh Riloff dan Thelen dalam bahasa Inggris. Bedanya, penelitian sebelumnya dikembangkan untuk dokumen berbahasa Arab sehingga hal ini tentunya akan berbeda dengan Bahasa Indonesia karena berpengaruh pada struktur kata-katanya.
- b. Kata tanya yang digunakan dalam query adalah kata tanya yang ada pada bahasa Indonesia yaitu: SIAPA, APA, DIMANA, KAPAN dan MENGAPA. Tipe pertanyaan BAGAIMANA, MANA dan akhiran –KAH tidak diikutsertakan dalam penelitian ini yang akan dijelaskan pada sub bab 3.5.
- c. Pola *query* pertanyaan yang diinputkan berupa kata pertama merupakan kata tanya dan selanjutnya merupakan pertanyaan yang diajukan dan diakhiri dengan kalimat tanya (?).
- d. Kamus yang digunakan dibuat secara manual oleh penulis.
- e. *Rules* yang digunakan dibuat berdasarkan *rules* yang dibuat oleh Riloff dan Thelen yang disesuaikan dengan bahasa Indonesia.
- f. Kalimat jawaban yang dikembalikan terhadap suatu *query* hanya satu kalimat atau tidak ada sama sekali.
- g. Pada evaluasi, kalimat jawaban akan dibandingkan dengan jawaban yang sebenarnya. Jawaban dianggap benar ketika kalimat yang dikeluarkan sistem sama dengan kalimat jawaban yang sebenarnya.

1.4. Tujuan

Tujuan yang dicapai dari penelitian ini adalah mengimplementasi *Question Answering System* dengan metode *rule-based* dalam dokumen berbahasa Indonesia, serta menganalisis pengaruh bahasa Indonesia pada implementasi *question answering system* untuk *reading comprehension test* dengan metode *rule-based*.

1.5. Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah yang akan digunakan adalah:

- a. Tahap studi literatur
Mengumpulkan referensi-referensi yang akan menunjang proses penelitian seperti buku, jurnal, paper, dan artikel dari internet tentang *Question Answering* dan *Rule-based method*.
- b. Tahap Pengolahan Data
Pada tahap ini, dilakukan pengambilan data dan *preprocessing data*. Data diambil dari situs detik.com dan selanjutnya data tersebut diubah sedemikian rupa agar bisa diolah oleh sistem. Preprocessing yang dilakukan meliputi *case folding*, *parsing* dan *stemming*. Data yang telah diolah selanjutnya diubah menjadi token-token agar dapat dibandingkan.
- c. Tahap Perancangan Sistem
Pada tahap ini, dilakukan perancangan sistem yang mampu menguji kemampuan dari metode *rule-based* pada kasus *Question Answering*.
- d. Tahap Implementasi
Pada tahap ini, dilakukan implementasi berdasarkan rancangan yang telah dibuat pada tahap sebelumnya. Implementasi menggunakan Python versi 3.5. Data yang berupa token-token kemudian dibandingkan agar dapat dinilai berdasarkan *rules* yang dibuat.
- e. Tahap Pengujian
Pada tahap ini, dilakukan pengujian terhadap sistem untuk mengetahui hasil implementasi yang telah dilakukan. Pengujian sendiri dicoba dengan 75 sampel pertanyaan dari 5 responden. Responden adalah mahasiswa Telkom University. Masing-masing responden memasukkan 10 sampel pertanyaan ke dalam sistem. Responden terlebih dahulu membaca dokumen yang dijadikan *dataset* oleh sistem. Dari hasil perbandingan antara hasil sistem dan hasil sebenarnya akan dihitung nilai *precision* dan *recall*.
- f. Tahap Analisis Hasil Pengujian
Pada tahap ini, dilakukan analisis dan penarikan kesimpulan berdasarkan hasil pengujian terhadap metode yang digunakan
- g. Tahap Pembuatan Laporan
Pada tahap ini, dilakukan penyusunan laporan dan pengumpulan dokumentasi berdasarkan hasil dari penelitian yang dilakukan.

1.6. Penyusunan Buku Tugas Akhir

Dalam penulisan Tugas Akhir ini akan dibagi menjadi beberapa bab yang meliputi hal-hal sebagai berikut:

- a. **BAB 1 : PENDAHULUAN**
Berisi latar belakang, perumusan masalah, tujuan, metodologi penyelesaian yang digunakan dalam penelitian tugas akhir, dan sistematika laporan tugas akhir ini.
- b. **BAB 2 : LANDASAN TEORI**
Bab ini memuat tentang teori-teori yang mendukung dalam perancangan sistem pengenalan angka tulisan tangan yang dibuat. Teori dasar yang dibahas antara lain tentang *Information Retrieval*, *Question Answering*, *Rule-based Method*, *Reading Comprehension*, *Parsing* dan *Stemming*.
- c. **BAB 3 : PERANCANGAN SISTEM**
Bab ini membahas tentang perencanaan perancangan sistem yang akan dibuat untuk membuat aplikasi "Metode *Rule-based Question Answering System* untuk *Reading Comprehension Test* dalam Bahasa Indonesia."
- d. **BAB 4 : PENGUJIAN DAN ANALISIS SISTEM**
Bab ini membahas tentang dokumentasi dan analisis terhadap pengujian implementasi dengan menggunakan data hasil *crawling* dari *detik.com* berupa artikel-artikel berita dari sistem "Metode *Rule-based Question Answering System* untuk *Reading Comprehension Test* dalam Bahasa Indonesia."
- e. **BAB 5 : KESIMPULAN DAN SARAN**
Bab terakhir ini berisi tentang kesimpulan hasil penelitian pada Tugas Akhir yang telah dilakukan dan saran terhadap pengembangan selanjutnya agar bisa penelitian ini dapat diteruskan.