1. Pendahuluan

1.1. Latar Belakang

Question Answering System dapat diartikan sebagai sistem pencarian informasi di mana jawaban diharapkan diperoleh atau direspon secara langsung terhadap pertanyaan yang diberikan dengan benar [1]. Salah satu jenis aplikasi *Question Answering System* adalah *Reading Comprehension* yaitu sebuah sistem yang memiliki data masukan berbasis pilihan ganda yang terdiri dari bacaan teks, pertanyaan, dan pilihan jawaban, dan kemudian dari data masukan tersebut maka sistem dapat menentukan jawaban mana yang benar [2]. Misalnya terdapat contoh *Reading Comprehension* dengan data masukan seperti yang ditunjukkan pada Gambar 1. Berdasarkan dengan bacaan teks, pertanyaan, dan pilihan jawaban yang ada maka akan diproses sedemikian rupa untuk mendapatkan jawaban yang benar.

P : Hari itu adalah ulang tahun Jessie si beruang. Dia mengadakan pesta. Dia meminta dua sahabatnya untuk datang ke pesta tersebut.Dia membuat kue besar dan menggantung beberapa balon.

Q : Siapa yang sedang ulang tahun ?

A :
A. Jessie si beruang
B. Tidak ada
C. Singa
D. Harimau

Gambar 1. Contoh dari sebuah Reading Comprehension yang telah diubah ke dalam Bahasa Indonesia

Dari hasil penelitian sebelumnya terdapat beberapa metode yang dapat digunakan untuk menyelesaikan sebuah *Reading Comprehension* seperti menggunakan *Bag of Word* [11], *Background Model* [12], dan *Baseline* [2]. Selain itu struktur teks yang dibandingkan untuk mendapatkan jawaban yang paling tepat juga bervariasi, seperti yang ditemukan dalam penelitian Stanovsky yang membandingkan penggunaan struktur teks asli, ketergantungan antara kata-kata, *Semantic Role Labelling* (SRL), *Open Information Extraction* [6]. Hasil penelitian Stanovsky menunjukkan bahwa penggunaan struktur *Open Informasi Extraction* menghasilkan kinerja tertinggi.

Open Information Extraction dapat diartikan sebagai metode yang digunakan untuk mengekstraksi informasi sebanyak mungkin tanpa harus menentukan jenis informasi yang akan diekstraksi [7]. Sistem *Open Information Extraction* akan memproses semua kalimat masukan dan mengekstraksi bagian-bagian penting dari kalimat masukan menjadi *tuple* relasi dalam *format* (*Argument, Relation, Argument*) yang mana *Relation* adalah objek yang memicu suatu hubungan dan *Argument* adalah objek yang memiliki ketergantungan pada suatu *Relation.* Berdasarkan penelitian sebelumnya telah dijelaskan bahwa *Open Information Extraction* memiliki kemampuan untuk merepresentasikan predikat multi-kata dan argumen yang memungkinkannya secara alami menggabungkan dua ide kesamaan. Konteks kata yang diperoleh dari *Open Information Extraction* yang sama dalam bentuk predikat atau argumen leksikal dapat dikatakan dekat dan menunjukkan kesamaan domain. Proses utama dalam sistem *Open Information Extraction* adalah melakukan ekstraksi relasi yang mengekstraksi *tuple* relasi dari kalimat masukan yang ada [6].

Terdapat beberapa metode yang dapat digunakan untuk melakukan ekstraksi relasi, yaitu pendekatan berbasis aturan untuk menemukan kesamaan teks dan metode berbasis pembelajaran [3]. Dalam pendekatan berbasis aturan, secara umum sistem akan menampilkan keluaran dalam bentuk serangkaian aturan pada domain tertentu untuk memecahkan suatu masalah [4], Berbeda dari pendekatan berbasis aturan, pendekatan berbasis pembelajaran biasanya akan menggunakan beberapa atau sekumpulan *dataset* untuk dilatih, kemudian hasil pelatihan tersebut akan digunakan untuk proses klasifikasi [5]. Dengan demikian dapat disimpulkan bahwa metode *Open Information Extraction* memiliki kinerja yang baik pada pasangan kata yang menunjukkan kesamaan topikal atau fungsional sehingga menjadikan metode *Open Information Extraction* cocok untuk menjadi metode ekstraksi informasi yang hasil ekstraksinya berupa *tuple* relasi tersebut dapat digunakan kembali pada proses pencocokan jawaban [6].

Dalam Tugas Akhir ini diusulkan sebuah metode untuk menyelesaikan sebuah *Reading Comprehension* dengan proses ekstraksi informasi menggunakan *Open Information Extraction* Bahasa Indonesia yang kemudian pada *tuple* relasi yang dihasilkan dari proses ekstraksi informasi dapat digunakan kembali pada proses pencocokan

jawaban. Dikarenakan saat ini tidak ada *dataset Reading Comprehension* dalam Bahasa Indonesia, kami menggunakan *dataset* terjemahan dari *MCTest* [2]. Sebagai bentuk evaluasi, pada sistem yang terpisah juga diterapkan metode *Baseline* yang di dalamnya akan langsung dilakukan proses pencocokan jawaban dari setiap data masukan yang diberikan dengan menggunakan algoritma *Sliding Windows* dan *Distance Based*, seperti yang telah dilakukan pada penelitian sebelumnya [2].

1.2. Topik dan Batasannya

Terdapat beberapa hal yang menjadi rumusan masalah dalam Tugas Akhir ini yaitu sebagai berikut :

1) Bagaimana tahapan untuk menyelesaikan sebuah *Reading Comprehension* dengan menggunakan keluaran berupa *tuple* relasi dari metode *Open Information Extraction* berbasis Bahasa Indonesia ?
2) Apakah penggunaan metode *Open Information Extraction* Bahasa Indonesia dapat meningkatkan performansi akurasi ketepatan jawaban pada *Reading Comprehension* jika dibandingkan dengan penggunaan metode *Baseline* ?

Selain itu sebagai data masukan pada sistem, telah digunakan 100 contoh pilihan ganda yang terdiri dari bacaan teks, pertanyaan dan pilihan jawaban dari *dataset MCTest* yang berasal dari penelitian Richardson et al. *Dataset MCTest* sendiri terbagi menjadi 3 jenis yaitu *MC160*, *MC500*, dan *MC660*. Kemudian, ketiga *dataset* tersebut juga kembali dibagi menjadi 3 bagian seperti *train*, *dev*, dan *test*. Dari semua *dataset* tersebut telah diambil 45 contoh pilihan ganda dari *dataset MC160 dev* dan 55 contoh pilihan ganda dari *dataset MC160 train*. *MCTest* sendiri adalah penelitian yang menyediakan serangkaian pilihan ganda berbasis bahasa Inggris yang dapat tersedia secara bebas yang dimaksudkan untuk bahan penelitian tentang *Reading Comprehension* [2]. Contoh pilihan ganda dari *dataset MCTest* yang digunakan pada Tugas Akhir ini tentunya juga memiliki batasan seperti :

1) Berbasis pilihan ganda sederhana yang terdiri dari bacaan teks, pertanyaan, dan serangkaian jawaban.
2) Pilihan ganda yang digunakan hanyalah pilihan ganda yang mana pertanyaan dan jawabannya tertera secara eksplisit pada bacaan teks atau jawaban dari pertanyaan itu haruslah memang sudah ada pada bacaan teks.

1.3. Tujuan

Tujuan dari adanya Tugas Akhir ini adalah sebagai berikut :

1) Menyelesaikan sebuah *Reading Comprehension* dengan menggunakan keluaran berupa *tuple* relasi dari metode *Open Information Extraction* berbasis Bahasa Indonesia.
2) Menemukan seberapa besar akurasi yang dihasilkan dari penggunaan metode *Open Information Extraction* berbasis Bahasa Indonesia pada sebuah *Reading Comprehension.*
3) Melakukan perbandingan kinerja dari penggunaan metode *Open Information Extraction* berbasis Bahasa Indonesia dengan metode *Baseline* pada sebuah *Reading Comprehension.*

1.4. Organisasi Tulisan

Kemudian selanjutnya masih terdapat beberapa Bab lainnya untuk memberikan penjelasan lebih lanjut mengenai apa saja yang dilakukan pada Tugas Akhir ini. Pada Bab 2 membahas mengenai studi terkait dari Tugas Akhir yang telah dikerjakan, pada Bab 3 membahas mengenai bagaimana sistem yang telah dibangun untuk dapat menyelesaikan sebuah *Reading Comprehension*, pada Bab 4 membahas mengenai evaluasi berupa analisis masalah dan hasil pengujian semua *dataset* yang ada terhadap sistem yang telah dibangun, terakhir pada Bab 5 akan dibahas mengenai kesimpulan.