

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

*Coreference resolution* adalah suatu proses identifikasi antara entitas pada sebuah wacana yang merujuk pada hal atau identitas yang sama[1]. *Coreference resolution* merupakan salah satu penelitian dalam bidang pemrosesan bahasa alami atau NLP (*Natural Language Processing*). Penelitian *coreference resolution* ini sudah banyak dilakukan dalam beberapa bahasa, seperti bahasa inggris dan bahasa cina. Pada penelitian *coreference resolution* bahasa indonesia, sudah dilakukan identifikasi entitas kata tunggal dan jamak berupa nama pada sebuah teks novel berbahasa indonesia.

Pada penelitian sebelumnya sudah dilakukan oleh Ayu Linggar Sari pada kasus kata ganti tunggal dan didapatkan nilai rata-rata akurasi sebesar 50.14%. Pada kasus entitas kata jamak oleh Syifa Muhammad Husni, didapatkan nilai rata-rata akurasi sebesar 61.41%[2].

Pendeteksian kata ganti pada suatu teks tidak hanya berupa entitas kata tunggal dan kata jamak seperti yang sudah diteliti oleh Ayu Linggar Sari dan Syifa Muhammad Husni. Penelitian terakhir oleh Syifa Muhammad Husni memiliki kekurangan yaitu belum adanya pendeteksian kata ganti berimbuhan seperti –NYA, -MU, dan –KU dikarenakan pada *preprocessing* tahap *stemming*, tahap tersebut menghilangkan setiap kata ganti imbuhan. Menurut buku “Tata Bahasa Praktis Bahasa Indonesia”, kata ganti dalam bahasa indonesia menurut peranannya juga memiliki kata ganti -NYA yang mana untuk menyatakan diri orang ketiga atau orang yang dibicarakan yang digunakan pada akhiran -NYA sebagai pengganti IA dalam fungsi objek atau sasaran dan dalam kontruksi kepemilikan[3]. Berikut adalah contoh kalimat yang mengandung *coreference* dengan imbuhan akhiran –NYA.

*Andi<sub>1</sub> hendak pergi berbelanja ke pasar. Namun, dia<sub>2</sub> melupakan dompetnya<sub>3</sub> yang tertinggal di rumah.*

Berdasarkan penjelasan di atas, belum ada pendeteksian kata ganti yang memiliki konstruksi kepemilikan -NYA. Selain pendeteksian kata ganti dari penelitian sebelumnya, penelitian ini akan menggunakan metode *Recurrent neural network* (RNN) dengan menggunakan arsitektur *many to one* yang dapat digunakan pada kasus *natural language processing* seperti pada penelitian Wiseman S., Rush A.M., Shieber S. M., digunakan untuk penalaran lebih dari berbagai sebutan menggunakan *coreference*[4].

Maka, penelitian ini bermaksud untuk mengembangkan sistem *coreference resolution* pada sebuah *dataset Indonesian Manually Tagged Corpus*

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang diuraikan, permasalahan yang dapat diidentifikasi adalah kurangnya nilai akurasi dari penelitian sebelumnya dan penambahan pendeteksian kata ganti kepemilikan.

## 1.3 Maksud dan Tujuan

Berdasarkan permasalahan yang diteliti, maka maksud dari penelitian ini adalah menambahkan identifikasi kata ganti kepemilikan pada *coreference resolution*. Sedangkan tujuan yang ingin dicapai adalah mengetahui nilai akurasi terbaru dari *Coreference resolution* pada kasus kata ganti kepemilikan menggunakan *Recurrent neural network*.

## 1.4 Batasan Masalah

Untuk menghindari pembahasan yang meluas, maka penelitian ini memiliki beberapa batasan masalah sebagai berikut:

### a. Masukan

Berikut adalah batasan masalah untuk data masukan:

- 1) Data yang digunakan berupa sebuah *dataset Indonesian Manually Tagged Corpus* dengan format .tsv.
- 2) Dokumen yang digunakan untuk bahan *training* merupakan *dataset Indonesian Manually Tagged Corpus*.
- 3) Dokumen yang digunakan untuk bahan *testing* merupakan *dataset Indonesian Manually Tagged Corpus*.

4) Isi teks yang ada pada *dataset Indonesian Manually Tagged Corpus* berupa konten berita (secara umum).

b. Proses

Berikut adalah batasan masalah untuk proses:

- 1) Menggunakan pelabelan *coreference*.
- 2) Tahap *Preprocessing* yang dilakukan adalah tokenisasi kalimat dan tokenisasi kata
- 3) Menggunakan ekstraksi fitur *i-propnoun* dan *j-propnoun* sebagai masukan untuk RNN

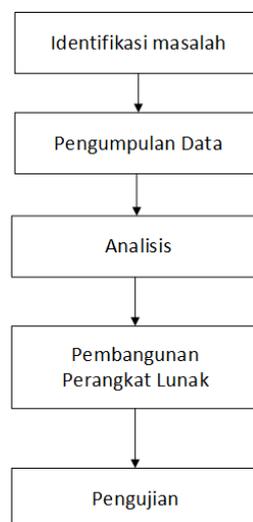
c. Keluaran

Berikut adalah batasan masalah untuk data keluaran:

- 1) Pasangan kata dengan hasil klasifikasi *coreference* atau bukan *coreference*.
- 2) Nilai akurasi dari hasil klasifikasi.

## 1.5 Metode Penelitian

Metode penelitian atau metode ilmiah adalah prosedur atau langkah-langkah sistematis dalam menyusun suatu penelitian[5]. Pada penelitian ini digunakan metode penelitian kuantitatif yang mana sifatnya dapat diklasifikasikan, kongkrit, teramati, dan terukur. Langkah-langkah Penelitian dapat dilihat pada Gambar 1.1 Skema Metode Penelitian Kuantitatif.



**Gambar 1.1 Skema Metode Penelitian Kuantitatif**

### 1.5.1 Identifikasi Masalah

Berdasarkan latar belakang yang diuraikan, permasalahan yang dapat diidentifikasi adalah kurangnya nilai akurasi dari penelitian sebelumnya dan penambahan pendeteksian kata ganti kepemilikan.

### 1.5.2 Pengumpulan Data

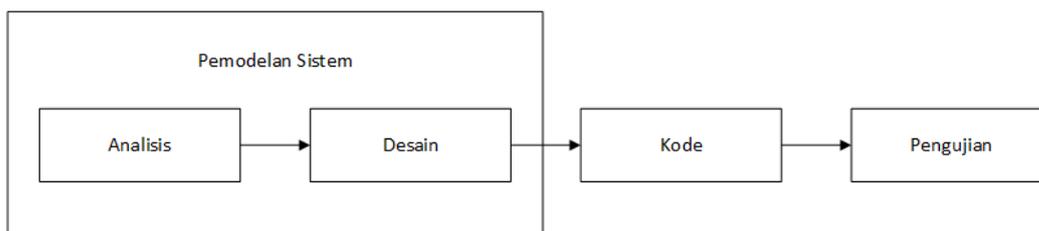
Pengumpulan data yang digunakan adalah dengan melakukan studi literatur yang dikumpulkan dalam bentuk tercetak maupun elektronik seperti buku, *e-book*, jurnal, paper, skripsi, dan beberapa sumber lainnya yang berhubungan dengan *coreference resolution*. Untuk data yang akan diproses menggunakan *dataset Indonesian Manually Tagged Corpus* dengan format *.tsv*.

### 1.5.3 Analisis

Tahap analisis yang dilakukan adalah analisis data masukan yang akan digunakan, analisis tahap *Preprocessing*, ekstraksi fitur, sampai analisis klasifikasi dengan menggunakan metode RNN.

### 1.5.4 Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan adalah *Sequential Linear Model*. Metode ini digunakan karena penelitian yang akan dilakukan diawali dengan tahap analisis dan pemodelan sistem yang cocok untuk membangun sistem *coreference resolution* dan diakhiri dengan tes atau pengujian. Tahapan-tahapan dari metode *Sequential Linear Model* dapat dilihat pada Gambar 1.2 *Sequential Linear Model*.



**Gambar 1.2** *Sequential Linear Model*

- a. Analisis Kebutuhan Sistem  
Analisis kebutuhan sistem terdiri dari penentuan perangkat lunak, penentuan perangkat keras, dan analisis data masukan yaitu *dataset Indonesian Manually Tagged Corpus* untuk mengetahui tahap *Preprocessing*.
- b. Desain sistem  
Langkah awal yang dilakukan adalah membuat rancangan sistem secara keseluruhan yaitu membuat skema *Preprocessing training*, dan *testing*.
- c. Implementasi atau kode  
Tahap ini merupakan tahap pembuatan aplikasi berupa pengkodean sesuai dengan desain yang telah dibuat pada tahap perancangan
- d. Pengujian  
Pengujian yang dilakukan meliputi metode *simple RNN* untuk mengetahui dampak akurasi terhadap *coreference resolution*.

### 1.5.5 Pengujian

Pengujian yang dilakukan dalam penelitian ini berupa pengujian sistem yang sudah dibangun berupa sebuah aplikasi dan pengujian akurasi.

### 1.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mengenai permasalahan dan pemecahannya. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

#### **BAB 1 PENDAHULUAN**

Bab ini membahas mengenai latar belakang, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian, serta sistematika penulisan untuk menjelaskan pokok-pokok pembahasannya.

#### **BAB 2 LANDASAN TEORI**

Pada bab ini akan menjelaskan mengenai objek dari penelitian, dan teori-teori pendukung yang berhubungan dengan pembangunan sistem

**BAB 3 ANALISIS DAN PERANCANGAN SISTEM**

Bab ini menganalisis masalah yang digambarkan pada bab 1 dengan lebih rinci seperti analisis masalah, analisis data masukan, analisis algoritma, dan analisis pembangunan aplikasi.

**BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM**

Bab ini berisi mengenai implementasi dan analisis dari perancangan sistem yang dilakukan, ujicoba dan hasil pengujian sistem

**BAB 5 KESIMPULAN DAN SARAN**

Bab ini berisi mengenai kesimpulan yang diperoleh dari hasil pengujian sistem serta saran untuk pengembangan sistem kedepan

