BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Pengenalan entitas bernama merupakan bagian penting hampir di semua proses pemprosesan bahasa alami seperti ekstraksi informasi, mesin penjawab otomatis, peringkasan, temu-kembali informasi, mesin penerjemah, anotasi video, *semantic web search*, dan biofarmatik. Pengenalan entitas bernama digunakan untuk mengidentifikasi dan mengklasifikasi entitas nama termasuk nama orang, lokasi, organisasi, waktu, dan kuantitas pada suatu teks. Namun pengenalan entitas bernama khususnya pada bahasa Indonesia masih jarang dilakukan. Hal ini akan mengakibatkan hasil sistem yang menggunakan pengenalan entitas bernama kurang maksimal seperti pada penelitian yang dilakukan Yusuf Syaifudin. Dalam penelitiannya masih terdapat kegagalan mengekstraksi kalimat pertanyaan yang disebabkan sistem pengenalan entitas bernama gagal mendeteksi entitas [1].

Salah satu cara melakukan pengenalan entitas adalah *machine learning*. Metode yang dapat digunakan dalam *machine learning* yaitu *Conditional Random Fields* karena terbukti efektif dan memberikan hasil yang baik dalam pelabelan sekuens data [2] . Metode CRF untuk kasus pengenalan entitas bernama berhasil diimplementasikan pada berbagai bahasa di dunia dan menghasilkan nilai akurasi yang tinggi seperti pada bahasa Inggris dengan akurasi 92,29% oleh Jenny Rose Finkel, Trond Grenager, dan Christopher Manning [3], bahasa Bengali akurasi mencapai 90,7% dilakukan oleh Asif Ekbal, Rejwanul Haque, Sivaji Bandyopadhyay [4]. Dari penelitian-penelitian yang telah dilakukan terlihat bahwa pengenalan entitas bernama menggunakan CRF memiliki nilai akurasi yang cukup tinggi. Namun, penelitian pengenalan entitas bernama menggunakan CRF masih belum diaplikasikan pada bahasa Indonesia sehingga masih diperlukan suatu penelitian untuk mengetahui keakuratan metode CRF dalam mengenali entitas bernama khususnya pada teks bahasa Indonesia.

Fitur yang digunakan adalah fitur kelas kata karena pada penelitian pengenalan entitas bernama menggunakan HMM menghasilkan akurasi yang

cukup baik yakni sebesar 88,4% [5]. Metode CRf diharapkan dapat mengenali entitas bernama khususnya pada teks bahasa Indonesia dan menghasilkab nilai akurasi yang baik sehingga dalam penelitian ini akan dilakukan "*Pengenalan Entitas Bernama pada Teks Bahasa Indonesia Menggunakan Conditional Random Fields*".

1.2. Rumusan Masalah

Berdasarkan penjelasan yang sudah diuraikan pada latar belakang, maka rumusan masalah dalam penelitian ini adalah seberapa baik metode CRF dapat mengenali entitas nama orang, lokasi, dan organisasi pada teks bahasa Indonesia.

1.3. Maksud dan Tujuan

Maksud dari penelitian ini adalah melakukan pengenalan entitas bernama menggunakan *Conditional Random Fields* untuk mengenali entitas nama orang, lokasi, dan organisasi pada teks bahasa Indonesia sedangkan tujuan dari penelitian ini adalah untuk mengetahui keakuratan metode CRF dalam mengenali entitas nama orang, lokasi, dan organisasi pada teks bahasa Indonesia.

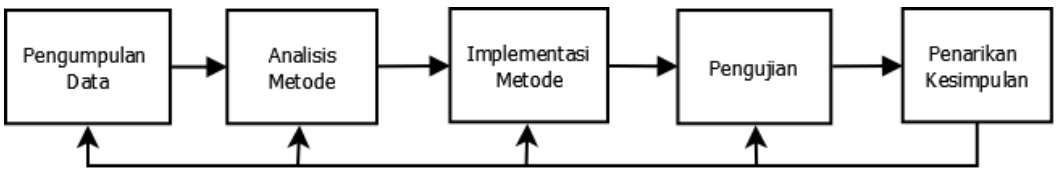
1.4. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Entitas yang dikenali adalah entitas nama orang, lokasi dan organisasi.
2. Fitur yang digunakan adalah fitur kelas kata.
3. Data masukkan berupa teks berita politik.

1.5. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini yaitu tahap pengumpulan data, analisis metode, implementasi metode, pengujian, dan penarikan kesimpulan yang dapat dilihat pada Gambar 1.1 berikut.



Gambar 1.1 Tahap Penelitian

Tahapan metodologi penelitian yang digunakan dalam penelitian tugas akhir ini adalah sebagai berikut :

1. Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini yakni studi literatur. Tahap ini dilakukan dengan cara mempelajari literatur-literatur seperti buku, jurnal, artikel ilmiah, dan *website* yang berhubungan dengan pengenalan entitas bernama dan c*onditional random field*. Tujuan dari studi literatur ini adalah sebagai dasar teori pembangunan simulator.

2. Analisis Metode

Pada tahapan ini dilakukan analisis kebutuhan dan analisis proses metode *conditional random fields* dalam mengenali entitas bernama. Setelah kebutuhan metode dan elemen simulator terkumpul maka dibuat perancangan simulator yang akan dibangun.

3. Implementasi Metode

Proses yang dilakukan dalam tahapan ini adalah membangun simulator pengenalan entitas berdasarkan analisis metode yang telah dilakukan. Simulator dibangun menggunakan bahasa pemograman java dan teks editor sebagai tempat menyimpan data pelatihan dan pengujian.

4. Pengujian

Setelah simulator dibangun tahap selanjutnya adalah melakukan pengujian hasil implementasi *conditional random fields* untuk mengenali entitas bernama pada teks bahasa Indonesia. Pengujian ini difokuskan pada pengujian akurasi data.

5. Penarikan kesimpulan

Di tahap ini dilakukan penarikan kesimpulan berdasarkan hasil pengujian implementasi metode *conditional random fields* untuk mengenali entitas bernama pada teks bahasa Indonesia.

1.6. Sistematika Penulisan

Sistematika penulisan laporan akhir penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Pada bab ini berisi penjelasan mengenai latar belakang, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, dan sistematika penulisan untuk memberikan gambaran urutan pemahaman dalam menyajikan laporan ini.

BAB 2 LANDASAN TEORI

Bab ini membahas landasan teori yang digunakan untuk menganalisis masalah dan mengolah data penelitian yaitu teori mengenai ekstraksi informasi, pengenalan entitas bernama, pendekatan pengenalan entitas bernama, metode CRF, pengenalan entitas bernama dengan CRF, akurasi, dan teori-teori pembangun simulator.

BAB 3 ANALISIS KEBUTUHAN ALGORITMA

Di bab ini diuraikan proses-proses mengenali entitas bernama pada teks bahasa Indonesia menggunakan metode CRF dan kebutuhan algoritma yang digunakan. Bab dijelaskan mengenai analisis dan perancangan simulator yang akan dibangun mulai dari analisis masalah, analisis data masukan, analisis metode CRF dalam mengenali entitas bernama pada teks bahasa Indonesia, analisis kebutuhan nonfungsional, pemodelan sistem, perancangan antarmuka, dan perancangan pesan.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab yang berisi penjelasan implementasi metode CRF dalam mengenali entitas bernama pada teks bahasa Indonesia dan pengujian akurasi simulator.

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil analisis metode CRF untuk mengenali entitas bernama pada teks bahasa Indonesia dan saran-saran untuk pengembangan pengenalan entitas bernama selanjutnya.