

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Name Entity Recognition (NER) atau Pengenalan Entitas Bernama adalah salah satu komponen utama dari *Information Extraction* yang bertujuan untuk mendeteksi dan mengklasifikasikan *named-entity* pada suatu teks. NER umumnya digunakan untuk mendeteksi nama orang, nama tempat dan organisasi dari sebuah dokumen, tetapi dapat juga diperluas untuk identifikasi gen, protein dan lainnya sesuai kebutuhan. NER juga merupakan komponen penting yang mendasari banyak aplikasi *Natural Language Processing* (NLP) seperti *information extraction* (IE), *Question Answering*, *Text Summarization*, dan *Information Retrieval*.

Named Entity Recognition untuk teks bahasa Indonesia sudah pernah diterapkan oleh Hadi Permana [1] menggunakan bi-LSTM dan mendapatkan akurasi dengan hasil terbaik untuk PERSON, ORGANIZATION, dan LOCATION ketiganya di bawah 50% dimana hasil tidak optimal ini dikarenakan minimnya fitur yang digunakan. Sedangkan ditemukan penelitian [2] dengan menggunakan *Conditional Random Field* (CRF) mendapatkan F₁ Score pada kelas PERSON, ORGANIZATION, dan LOCATION sebesar 72%, 66%, dan 84%. Sehingga menjadikan penelitian dengan metode CRF [2] sebagai *state-of-the-art* pada kasus NER bahasa Indonesia.

Pada penelitian yang dilakukan oleh Li et al [3] tentang *Biomedical Named Entity Recognition* pada teks bahasa Inggris dari BioCreative II GM dan ditemukan bahwa metode *Elman Recurrent Neural Network* (ERNN) mengungguli metode CRF untuk menentukan gen atau protein. Pada penelitian yang dilakukan oleh Sun et al [4] ditemukan bahwa Modified ERNN mengungguli CRF pada kasus *Named Entity Recognition* untuk teks bahasa Inggris. Tetapi pada kasus teks bahasa Indonesia ERNN masih belum diterapkan sehingga tidak dapat diketahui performanya

Berdasarkan uraian tersebut maka pada penelitian ini akan dilakukan implementasi metode *Elman Recurrent Neural Network* pada kasus *Named Entity Recognition* untuk teks bahasa Indonesia.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang dipaparkan di atas, masalah yang teridentifikasi belum adanya penelitian yang menerapkan ERNN pada kasus *Named Entity Recognition* untuk teks bahasa Indonesia sehingga belum diketahuinya performa yang dihasilkan.

1.3 Maksud dan Tujuan

Maksud dari Penelitian ini adalah untuk menerapkan metode ERNN pada kasus *Named Entity Recognition* untuk teks bahasa Indonesia. Sedangkan tujuan dari penelitian ini adalah untuk mengukur performa berdasarkan F_1 Score dan Akurasi untuk metode ERNN pada kasus *Named Entity Recognition* untuk teks bahasa Indonesia.

1.4 Batasan Masalah

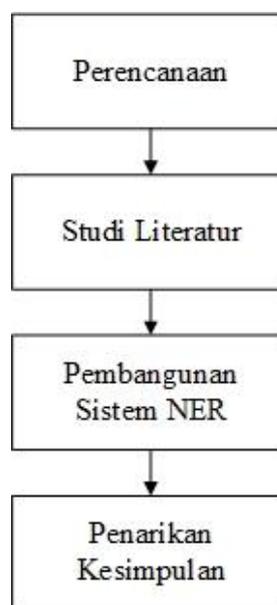
Peneliti membatasi permasalahan sebagai berikut.

1. Masukan
 - a. Format file dari data masukan yang digunakan adalah *.txt.
 - b. Data masukan untuk Training berupa Annotated Corpus dan data masukan untuk Testing berupa Teks berita berbahasa Indonesia.
 - c. Corpus yang digunakan berasal dari penelitian M. Fachri [5] dan Rusliani [6].
 - d. Dataset yang digunakan bertipe teks dan diambil dari berita politik.
2. Proses
 - a. Fitur yang digunakan dalam penelitian ini adalah Kata dan *Spelling Features*.

- b. Terdapat empat jenis *Spelling Features* yaitu *First Capital*, *All Capital*, *All Lower*, dan *All Digit*.
 - c. Terdapat enam kelas entitas yang dikenali pada penelitian ini yaitu nama orang, lokasi, organisasi, kuantitas, dan waktu.
3. Keluaran
 - a. Data keluaran yang dihasilkan oleh sistem berupa label kelas entitas.

1.5 Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode kuantitatif [8]. Metode ini digunakan karena data yang digunakan dalam penelitian ini berbentuk angka serta bisa diukur secara akurat dengan alat yang objektif. Alur Penelitian yang digunakan dapat dilihat pada Gambar 1.1.



Gambar 1. 1 Alur Penelitian

1.5.1 Perencanaan

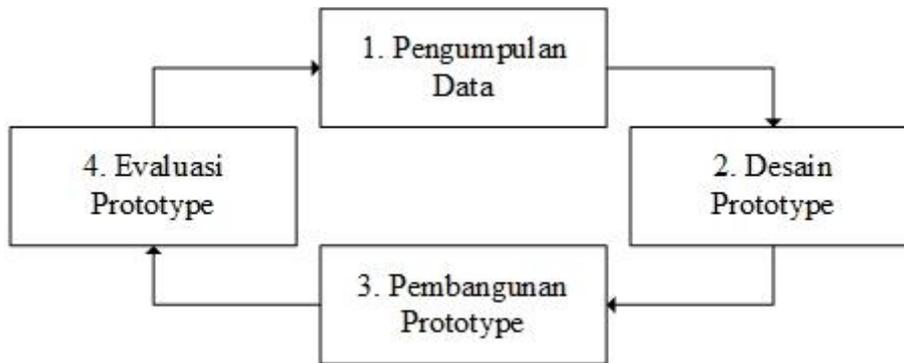
Tahap perencanaan dalam penelitian ini memiliki beberapa tahapan yaitu, tahapan pertama adalah identifikasi masalah, tahap selanjutnya adalah menetapkan tujuan penelitian dan tahapan terakhir adalah menentukan data yang diperlukan.

1.5.2 Studi Literatur

Studi Literatur yang dilakukan pada penelitian ini adalah melakukan studi pustaka, informasi terkait penelitian tentang Implementasi ERNN pada Kasus *Named Entity Recognition*.

1.5.3 Pembangunan Sistem NER

Metode pembangunan perangkat lunak yang digunakan adalah metode *prototype* [7]. Metode *prototype* dipilih karena pada penelitian memiliki tujuan menguji performa dari ERNN pada NER bahasa Indonesia sehingga perlu dilakukannya evaluasi. Metode *prototype* dapat dilihat pada Gambar 1.2.



Gambar 1. 2 Metode *Prototype*

1) Pengumpulan Data

Pengumpulan kebutuhan merupakan tahap mengumpulkan data-data yang dibutuhkan dengan cara mengumpulkan literatur-literatur yang berkaitan dengan penelitian mengenai *Named Entity Recognition* dan *Elman Recurrent Neural Network*.

2) Desain *Prototype*

Pada tahap ini dilakukan perancangan pembuatan perangkat lunak. Proses ini berfokus pada pembuatan DFD, Diagram Konteks, dan perancangan antarmuka.

3) Pembangunan *Prototype*

Pada tahap ini dilakukan pembangunan *prototype* meliputi preprocessing, proses pelatihan (training), dan proses pengujian (testing) dengan menggunakan metode ERNN dalam pembangunan program.

4) Evaluasi *Prototype*

Pada tahap ini dilakukan pengujian dengan mengubah parameter untuk melihat performa ERNN.

1.5.4 Penarikan Kesimpulan

Pada tahap ini akan menyimpulkan hasil dari pembangunan Sistem NER yang dilakukan.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan penelitian ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian serta sistematika penulisan yang dimaksudkan agar dapat memberikan gambaran tentang penelitian ini dalam menyajikan penelitian mengenai *Named Entity Recognition* dengan menggunakan metode *Elman Recurrent Neural Network*.

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai landasan teori yang dipakai dalam data penelitian ini yaitu teori mengenai bahasa Indonesia, *Named Entity Recognition*, metode *Elman Recurrent Neural Network*, Matriks, dan bahasa pemrograman.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini berisi penjelasan secara rinci dari analisis dan perancangan penelitian ini dimulai dari analisis masalah, analisis solusi, analisis data masukan, analisis proses,

analisis kebutuhan non fungsional, analisis kebutuhan fungsional, perancangan antarmuka, perancangan jaringan semantik, dan perancangan prosedural.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang pengimplementasian analisis dan perancangan yang telah dilakukan pada bab sebelumnya. Implementasi yang dilakukan terdiri dari implementasi perangkat keras, implementasi perangkat lunak, dan implementasi antarmuka. Pengujian yang dilakukan pada bab ini adalah pengujian Fungsionalitas, pengujian Parameter, dan Pengujian Performa.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi tentang hasil yang diperoleh dari pengujian yang dilakukan pada bab sebelumnya. Bab ini juga memberikan tambahan-tambahan yang bisa membangun penelitian selanjutnya.