**Pendekatan Conditional Random Field dan POS-Tagging pada Indonesian Named Entity Recognition**

Proposal Proyek Pemrosesan Bahasa Alami

Oleh :

| 11S18022 | Devi Wahyuni Silitonga |
| --- | --- |
| 11S18037 | Hisar Haryanto Sinaga |
| 11S18046 | Simson Fransisco Panjaitan |
| 11S18058 | Leonardo Sianturi |
| 11S18066 | Jumadi Heryanto Damanik |



Institut Teknologi Del

2020/2021

# DAFTAR ISI

[**DAFTAR ISI**](#_heading=h.mcci74q52csp) **1**

[**PENDAHULUAN**](#_heading=h.3znysh7) **2**

[**Latar Belakang**](#_heading=h.2et92p0) **2**

[**Alasan Pemilihan Studi Kasus**](#_heading=h.tyjcwt) **4**

[**Alasan Pemilihan Pendekatan**](#_heading=h.mvmw6cjpatyt) **4**

[**Ruang lingkup**](#_heading=h.4d34og8) **4**

[**Hasil yang diharapkan**](#_heading=h.fv2u1my4a7r4) **5**

# PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang topik penelitian, rumusan permasalahan penelitian, tujuan penelitian, ruang lingkup penelitian, metode penelitian dan sistematika penulisan dalam menyusun Tugas Akhir ini.

## Latar Belakang

Dokumen adalah hal penting di era teknologi saat ini. Dokumen dapat diperoleh dalam bentuk online maupun offline. Melalui dokumen, terdapat banyak informasi penting yang dapat diambil dari setiap kalimatnya, seperti nama orang, ataupun nama suatu tempat atau wilayah. Informasi di dalam dokumen dapat diperoleh melalui membaca dokumen tersebut secara keseluruhan, dan apabila dokumen berisi teks yang sangat panjang, maka dibutuhkan waktu yang lama untuk mendapatkan informasi yang disajikan di dalam dokumen tersebut. Untuk mengatasi masalah tersebut, diciptakan lah NER (Named Entity Recognition) yang dapat digunakan untuk memperoleh informasi secara otomatis dari sebuah dokumen teks melalui proses ekstraksi dan klasifikasi kata atau frasa secara otomatis kedalam kategori tertentu.

Pada penerapannya, dapat dilakukan metode POS-Tagging untuk melakukan NER, dimana akan dilakukan tagging pada setiap entity. Proses tagging dapat dilakukan dengan cara Automatic Tagging, dimana proses tagging dilakukan pada setiap kata atau frasa sesuai dengan jenis entitasnya. Namun, terkadang terjadi kesalahan saat melakukan tag entitas, dimana kata atau frasa yang seharusnya memiliki entitas tidak diberikan tag. Maka dari itu, dilakukan pengimplementasian POS-Tagging yang bertujuan untuk memperoleh seluruh kata atau frasa yang memiliki kemungkinan mempunyai jenis entitas, selanjutnya akan dilakukan pengecekan terhadap kata atau frasa tersebut dengan rule yang telah ditentukan kemudian diberikan tag pada masing-masing entitas berdasarkan aturan yang telah ditentukan.

## Alasan Pemilihan Studi Kasus

*Indonesian Named-Entity Recognition* merupakan model yang dapat mengekstraksi informasi seperti nama orang, lokasi, maupun nama organisasi dari suatu kalimat atau teks yang banyak. Dengan adanya model *Indonesian Named-Entity Recognition* ini, dapat membantu pengguna jika pengguna ingin menemukan nama orang, lokasi, maupun nama organisasi dari suatu teks yang banyak dalam waktu yang singkat. Dengan demikian, berdasarkan hasil diskusi kelompok, kelompok sepakat untuk memilih studi kasus yang kedua karena studi kasus ini lebih unik dan menarik untuk dikerjakan dibandingkan studi kasus satu dan tiga.

## 

## Alasan Pemilihan Pendekatan

Pada tahun 2015, penelitian tentang Named Entity Recognition yang dilakukan oleh Lonandya et al menggunakan algoritma Semi-Supervised learning, proses automatic tagging memiliki kekurangan, dimana proses tagging pada setiap kata atau frasa tidak berjalan dengan baik. Pada penelitian tersebut, proses automatic tagging tidak dapat melakukan tag entitas pada kata atau frasa yang seharusnya tidak memiliki entitas. Untuk meningkatkan kemampuan automatic tagging nya, diterapkan POS-Tagging dengan beberapa aturan tertentu pada proses automatic tagging-nya.

## Ruang lingkup

Adapun ruang lingkup dari penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya dapat menerima inputan berupa kalimat atau teks berbahasa Indonesia yang sesuai dengan Ejaan Bahasa Indonesia.
2. Penelitian ini hanya mengklasifikasikan entitas informasi nama orang, nama lokasi, dan nama organisasi.

## Hasil yang diharapkan

Adapun hasil yang diharapkan dari penelitian ini adalah berupa model *Indonesian Named-Entity Recognition* yang dapat digunakan untuk melakukan tag entitas dan pengujian pada kalimat atau teks berbahasa Indonesia.