BAB I
PENDAHULUAN

Bab pertama di buku ini akan memberi penjelasan mengenai tugas akhir "Nested Named-Entity Recognition dalam Bahasa Indonesia Menggunakan Sequence-to-Set Network" secara keseluruhan, seperti latar belakang dan tujuan dari penyusunan tugas akhir ini, batasan penelitian yang dilakukan, dan juga dijelaskan sistematika pembahasan pelaksanaan tugas akhir di buku ini.

1.1 Latar Belakang

Cabang ilmu komputer yang mempelajari bagaimana caranya computer dapat memahami dan menganalisa bahasa manusia adalah cabang ilmu Pengolahan Bahasa Alami, umumnya dikenal sebagai *Natural Language Processing* (NLP). Ilmu ini memiliki kegunaan yang luas, seperti memahami bahasa manusia meskipun ada beragam bahasa didunia ini. Dengan memahami arti dari kalimat yang diberikan, computer bisa melakukan beragam *task*. *Task* yang saat ini sudah ditemukan untuk dilakukan seperti asisten virtual dalam *handphone* yang paling umum, bisa melakukan perintah untuk telepon kontak, memainkan musik, dan sebagainya. Bisa juga menerima kata – kata yang diucapkan dan diketik secara otomatis oleh komputer. Komputer tidak hanya dapat mengerti secara makna kalimat, tetapi juga bisa mengenal kata – kata sebagai entitas sendiri. Mengenal dalam suatu kalimat bagian apa yang merupakan subyek, obyek, suatu predikat dan sebagainya. Entitas-entitas dalam bahasa memiliki juga bagian sendiri yaitu entitas bernama, contoh kasus entitas bernama adalah entitas seseorang direpresentasikan nama seseorang, bisa juga suatu lokasi seperti rumah sakit, disebutkan nama dan rumah sakit tersebut.

Namun cabang ilmu NLP masih berkembang dalam bahasa Indonesia. Tidak asing bahwa sebagian besar metode yang digunakan terinspirasi dari metode NLP bahasa Inggris, karena memang bidang NLP untuk perkembangannya yang memimpin adalah bahasa Inggris sendiri. Salah satu contoh penggunaan metode

dari bidang NLP yang sering dilakukan penelitian adalah pengenalan entitas bernama, lebih umum disebut *Named Entity Recognition* (NER). *Task* ini berguna untuk mengenal entitas dalam kalimat untuk menemukan informasi yang dapat diolah lebih jauh lagi oleh komputer. Kegunaan mengenal entitas – entitas bernama dalam satu kalimat ini dapat ditemukan dalam teknologi terkini seperti *engine* untuk pencarian dan rekomendasi lebih optimal dan cepat karena NER, atau otomasi penentuan kategori tiket layanan *customer service*[1]. Namun, dari *task* NER ada beberapa kesusahan yang ditemukan, salah satunya adalah pengenalan entitas bernama tetapi yang bersarang. Entitas Bernama yang bersarang adalah penemuan yang baru ditemukan beberapa tahun lalu, dan telah dilakukan beberapa penelitian dengan beberapa metode. Tetapi kekurangan dari penelitian tersebut bukan pada akurasi (karena yang didapat adalah akurasi yang tinggi) tetapi tidak melihat sisi komputasi dalam persiapan dan training model. Kesulitan ini terjadi di penelitian sebelum – sebelumnya tentang permasalahan entitas bernama bersarang seperti metode *sequence-to-sequence*[2] juga metode *span-based*[3]. Tugas akhir ini menggabungkan manfaat metode Sequence-to-Set yang melawan metode span-based dan juga manfaat mendalami penelitian mengenai NER, khususnya yang bersarang, dalam bahasa Indonesia.

Dalam tugas akhir ini, penulis akan melakukan pengenalan entitas yang bersarang yang menangkal kekurangan yang disebut dengan metode Sequence-to-Set Network. Pendekatan metode ini akan menggunakan pendekatan *supervised*, dimana metode akan menggunakan data yang sudah dianotasi untuk proses *training* dan *testing*. Bagian model untuk proses *encoding* akan mengggunakan Bidirectional Encoder Representations from Transformers (BERT) dan Bi-

---

[1] Christopher Marshall, "What is named entity recognition (NER) and how can I use it?" (https://medium.com/mysuperai/what-is-named-entity-recognition-ner-and-how-can-i-use-it-2b68cf6f545d)

[2] Jana Strakova, Milan Straka, Jan Hajic. "*Neural architectures for nested ner through linearization*", Proceedings of ACL 2019, 2019.

[3] Mohammad Golam Sohrab and Makoto Miwa. "*Deep exhaustive model for nested named entity recognition*", Proceedings of EMNLP 2018, 2018.

directional Long Short Term Memory (BiLSTM), dan untuk bagian *decoder* nya akan menggunakan *non-autoregressive decoder* yang menggunakan *self-attention* juga *cross-attention* untuk mendapatkan ketergantungan antar entitas. Diharapkan hasil dari penelitian ini, yaitu pengenalan entitas yang bersarang dapat digunakan untuk penelitian linguistik komputasi bahasa Indonesia kedepannya.

1.2 Tujuan

Terdapat beberapa tujuan yang akan dicapai dari penelitian ini yang diharapkan oleh penulis. Bab ini menjelaskan beberapa tujuan yang akan dicapai. Beberapa tujuan tersebut terdiri dari:

- Melakukan pengenalan entitas, baik yang bersarang dan tidak, dari setiap kata dalam kalimat.
- Pembuatan program untuk pengenalan entity bersarang dalam bahasa Indonesia untuk membantu aplikasi dan penelitian NLP kedepannya.

1.3 Batasan Penelitian

Pada bagian ini akan batasan untuk menjelaskan bagian yang akan dikerjakan pada tugas akhir ini. Batasan ini adalah hal-hal ditentukan dalam penelitian tugas akhir ini yang tidak akan dilakukan oleh sistem. Berikut adalah beberapa batasan yang dimiliki pada tugas akhir ini :

1) Program ini hanya menerima input dalam Bahasa Indonesia.
2) Program akan berjalan secara offline.
3) Entitas yang diidentifikasikan sebanyak tujuh yaitu person, organization, date, time, event, location, miscellaneous.
4) Input kalimat hanya dalam 2 - 3 kalimat.
5) Dataset yang digunakan berasal dari tugas akhir Georgia Nikita (218116685)[4], dimana dataset bersumber dari berita dari situs berita CNN Indonesia dan telah dilabel kembali dan ditambahkan pelabelan NER secara bersarang dari dataset

[4] Georgia Nikita, Skripsi: "Service Oriented Nested NER untuk Ekstraksi Keyword Entitas di Portal Berita Bahasa Indonesia" (Surabaya: 2022).

tugas akhir Christian Nathaniel Purwanto (214116299)[5] dan Amelinda Tjandra Dewi (214116288)[6].

6) Representasi kata Part-of-Speech Tagging tidak digunakan.
7) Jika ada, akan dilakukan penyesuaian terhadap metode yang digunakan dalam menyelesaikan permasalahan pada tugas akhir ini.
8) Hasil akurasi (F1 Score) dari model yang dibuat memiliki target melebihi akurasi metode perbandingan.

1.4 Sistematika Pembahasan

Dalam subbab ini akan dijelaskan garis besar isi dari setiap bab yang ada pada buku tugas akhir ini. Sistematika ini akan membantu pembaca untuk mengetahui struktur dari pembahasan yang dirangkai penulis untuk memudahkan penjelasan penelitian yang dilakukan. Berikut adalah sistematika pembahasan yang dibuat untuk memudahkan pemahaman isi dari setiap bab secara garis besar.

- BAB I : PENDAHULUAN
  Dalam bab ini akan dijelaskan mengenai latar belakang, tujuan, ruang lingkup, batasan penelitian dari tugas akhir buku ini.
- BAB II : TEORI PENUNJANG
  Teori penunjang adalah bagian bab yang menjelaskan teori - teori yang menjadi rujukkan informasi yang digunakan dalam pembuatan dann penjelasan penelitian tugas akhir.
- BAB III : ARSITEKTUR SISTEM

[5] Christian Nathaniel Puerwono, Skripsi: Ekstraksi Entity dan Relasi Dalam Bahasa Indonesia Menggunakan Bidirectional LSTM" (Surabaya: 2018).

[6] Amelinda Tjandra Dewi, Skripsi:Named Entity Recognition dan Coreference Resolution Nama Orang untuk Teks Bahasa Indonesia dengan Menggunakan Conditional Random Fields. (Surabaya:2018).

Pada bab arsitektur sistem akan menjelaskan alur sistem yang dilewatkan tugas akhir ini. Penjelasan akan mengandung arsitektur secara keseluruhan dan juga untuk tiap bagian arsitektur umum tersebut.

- BAB IV : NESTED NER DALAM BAHASA INDONESIA

  Bab ini membawakan penjelasan mengenai Nested NER yang dibutuhkan untuk mengetahui teori, informasi, manfaat dan penggunaan Nested NER dalam bahasa Indonesia khususnya. Dataset dan juga proses *preprocessing* akan dibahas secara mendalam dibab ini sebelum memasuki pembahasan program / metode.

- BAB V : SEQUENCE TO SET DALAM BAHASA INDONESIA

  Pada bab kelima dijelaskan mengenai teori, cara kerja, dari metode Sequence to Set tersebut. Tiap bagian dari struktur metode akan dibahas secara rinci, juga dalam bab ini akan menjelaskan modifikasi yang dilakukan untuk tugas akhir ini. Juga ada *tracing* dalam contoh kasus untuk membantu penjelasan cara kerjanya metode.

- BAB VI : UJI COBA

  Dalam bab ini akan menjelaskan langkah-langkah yang dilakukan dalam melakukan uji coba serta hasil dari uji coba yang telah dilakukan. Juga uji coba untuk metode perbandingan akan dilakukan dan juga diberikan kesimpulan perbandingannya.

- BAB VII : PENUTUP

  Dalam bab ini akan membahas kesimpulan dari tugas akhir dan saran bagi pembaca buku tugas akhir.