BAB V

# SEQUENCE TO SET DALAM BAHASA INDONESIA

Pada bab ini akan dibahas mengenai pendahuluan mengenai NER, juga teori nested NER juga contoh-contoh penggunaan tugas NER dalam kehidupan sehari-hari. Selain itu, dataset yang digunakan dalam tugas akhir ini juga akan dibahas dengan detail. Penjelasan baik untuk dataset bahasa Inggris yang utama digunakan dari penelitian, maupun dataset bahasa Indonesia yang digunakan dalam penelitian ini. Dataset akan dibahas struktur/bentuk, visualisasi dan juga jenis tagset/jenis entitas yang ditentukan dalam tiap dataset. Bab ini juga ada subbab bagian pra proses (*preprocessing*) dengan rinci untuk mengetahui apa saja yang perlu dimodifikasi dari dataset mentah menjadi dataset yang akhir agar dapat diterima untuk training model.

## Named Enity Recognition (NER)

Sebelum mencoba mengerti apa itu tugas pengenalan named entity (atau NER) secara mendalam, perlu diketahui terlebih dahulu apa yang dapat disebut sebagai sebuah named entity. Istilah named entity awalnya dianggap memiliki relasi dekat dengan pembahasan mengenai *rigid designators* oleh Kripke. Namun diskusi named entity mengarah kepada rigid designators menjadi terlalu filosofis. Sehingga dalam penelitian dicarikan penjelasan yang lebih jelas dan ringkas. Sampai saat ini belum ada persetujuan yang resmi dari bidang NLP mengenai definisi resmi NER. Tetapi dalam skenario penelitian NER dapat disimpulkan definisi umum untuk tugas NER maupun arti sebuah named entities ini. Oleh pihak CoNLL 2002 sendiri, named entities adalah frase yang mengandung nama oleh seseorang, suatu organisasi, sebuah lokasi, sebuah waktu dan sejumlah kuantitas.