

**Membangun sebuah Combinatory Categorical
Grammar (CCG) Supertagger berbasis
Maximum Entropy untuk Bahasa Indonesia**

Proposal Tugas Akhir

Kelas TA NLP

Wisnu Adi Nurcahyo

NIM: 1301160479



Program Studi Sarjana Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2019

Lembar Persetujuan

**Membangun sebuah Combinatory Categorical Grammar (CCG)
Supertagger berbasis Maximum Entropy untuk Bahasa Indonesia**

***Building a Maximum Entropy based Combinatory Categorical
Grammar (CCG) Supertagger for Bahasa Indonesia***

Wisnu Adi Nurcahyo

NIM: 1301160479

Proposal ini diajukan sebagai usulan pembuatan tugas akhir pada
Program Studi Sarjana Informatika
Fakultas Informatika Universitas Telkom

Bandung, 24 Oktober 2019
Menyetujui

Calon Pembimbing 1

Dr. Ade Romadhony, S.T., M.T.
NIP: 06840042

Abstrak

Riset pemrosesan bahasa natural untuk bahasa Indonesia saat ini terbilang sedikit. Bahkan, masih banyak area riset yang belum tersentuh seperti contohnya *combinatory categorial grammar* (CCG). CCG merupakan formalisme tata bahasa yang pada akhirnya dapat dimanfaatkan untuk memperoleh informasi dari suatu kalimat. Informasi tersebut diperoleh setelah melakukan *parsing* berdasarkan formalisme CCG dengan menggunakan CCG *parser*. Untuk dapat melakukan *parsing*, CCG *parser* membutuhkan CCG *lexicon* yang mengandung bentuk formal dari suatu token kata. Bentuk formal tersebut umumnya adalah *combinatory logic*. CCG *lexicon* diperoleh dari proses pelabelan suatu token kata terhadap bentuk formalnya dengan menggunakan *supertagging*. Proses *supertagging* akan menghasilkan *supertag* yang kemudian disebut sebagai CCG *lexicon* karena formalisme yang digunakan adalah formalisme CCG.

Tugas akhir dengan judul **Membangun sebuah Combinatory Categorical Grammar (CCG) Supertagger berbasis Maximum Entropy untuk Bahasa Indonesia** berusaha untuk membangun versi awal dari CCG *supertagger* untuk bahasa Indonesia dengan harapan dapat menjadi inisiator riset pemrosesan bahasa natural dengan tema CCG sehingga ke depannya akan ada lebih banyak riset mengenai CCG yang tersedia. *Supertagger* tersebut akan dibangun dengan menggunakan model Maximum Entropy dan implementasinya akan ditulis dalam bahasa pemrograman Haskell.

Kata Kunci: combinatory categorial grammar, supertagger, maximum entropy model, bahasa indonesia, haskell

Daftar Isi

Abstrak	i
Daftar Isi	ii
I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	1
1.3 Tujuan	2
Daftar Pustaka	3
Lampiran	4

Bab I

Pendahuluan

1.1 Latar Belakang

Riset pemrosesan bahasa natural untuk bahasa Indonesia saat ini terbilang sedikit. Bahkan, masih banyak area riset yang belum tersentuh seperti contohnya *combinatory categorial grammar* (CCG). CCG merupakan sebuah formalisme tatabahasa yang dapat dimanfaatkan untuk membangun CCG *parser*. CCG *parser* dapat digunakan untuk mendapatkan berbagai macam informasi dari suatu kalimat. Sebagai contoh, CCG *parser* dapat mem-*parse* kalimat "sebutkan negara-negara yang bertetangga dengan Indonesia" ke dalam bentuk formal yang dapat dipahami oleh komputer yaitu $\lambda(x).negara(x) \wedge bertetangga(x, Indonesia)$. Bentuk formal tersebut merupakan *lambda calculus* (umumnya bentuk formal yang digunakan adalah *combinatory logic*) yang kemudian dapat diproses oleh komputer strukturnya dan akan mendapatkan *query* untuk mencari x yang berupa suatu negara dan memiliki ketetanggaan dengan Indonesia.

CCG *parser* membutuhkan CCG *lexicon* (atau dikenal juga sebagai CCG *supertag*) untuk dapat melakukan tugasnya. Sejauh ini belum ditemukan adanya riset mengenai CCG untuk bahasa Indonesia termasuk pada tahap awalnya yaitu pembentukan CCG *supertag*. Demikian itu, tugas akhir ini diharapkan dapat menjadi inisiator riset untuk area CCG dalam pemrosesan bahasa alami yaitu dengan membangun perangkat lunak untuk menghasilkan CCG *supertag* bahasa Indonesia. Proses yang menghasilkan *supertag* tersebut bernama *supertagging* adapun perangkat lunaknya bernama *supertagger*.

1.2 Perumusan Masalah

Berikut adalah rumusan masalah yang akan diangkat:

1. Mengapa CCG *supertagger* diperlukan?
2. Apa saja yang harus dipersiapkan untuk membangun CCG *supertagger*?
3. Bagaimana proses membangun CCG *supertagger*?

1.3 Tujuan

Berikut adalah tujuan yang diharapkan dapat dicapai oleh tugas akhir ini:

1. Mengenalkan alternatif metode yang dapat digunakan dalam pemrosesan bahasa alami untuk bahasa Indonesia.
2. Merilis CCG *supertagger* pertama untuk bahasa Indonesia.
3. Membuka peluang riset untuk CCG *parser* bahasa Indonesia.

Daftar Pustaka

Lampiran