CCGtown: Alat Anotasi Combinatory Categorial Grammar (CCG) Semi-otomatis Berbasis Web

Wisnu Adi Nurcahyo[1], Ade Romadhony[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]nurcahyo@student.telkomuniversity.ac.id, [2]aderomadhony@telkomuniversity.ac.id

Alat anotasi yang tersedia untuk *combinatory categorial grammar* (CCG) masih sangat terbatas. Salah satu aplikasi dengan antarmuka grafis yang dapat digunakan adalah CCGweb. Untuk dapat menggunakan CCGweb, pengguna diharuskan untuk melakukan *setup* secara mandiri. Proses ini dapat menjadi keterbatasan bagi *annotator* karena *dependency* yang perlu dipersiapkan cukup banyak. Selain itu, CCGweb hanya dapat mengerjakan satu proyek anotasi CCG dalam satu waktu. Demikian itu, kami membangun alat anotasi CCG yang dapat langsung digunakan oleh pengguna tanpa perlu melakukan *setup* terlebih dahulu. Alat anotasi tersebut kami beri nama CCGtown. Dengan CCGtown, pengguna dapat membuat banyak proyek anotasi sekaligus dan telah dilengkapi fitur yang sebagian besarnya tersedia di CCGweb. CCGtown merupakan alat anotasi CCG alternatif yang dapat digunakan oleh *annotator* CCG.

Kata kunci: pemrosesan bahasa alami, combinatory categorial grammar, alat anotasi

Abstract
Annotation tools available for combinatory categorial grammar (CCG) are still limited. One of the annotation tool with a graphical interface that can be used is CCGweb. To be able to use CCGweb, users are required to setup independently. This process can be a limitation for annotators because of the many dependencies that need to be prepared. CCGweb also can only work on one CCG annotation project at a time. With that in mind, we built a CCG annotation tool that users can use right away without any setup required. We call the annotation tool CCGtown. With CCGtown, users can create multiple annotation projects at once and it has features that are mostly available on CCGweb. CCGtown is an alternative CCG annotation tool which can be used by CCG annotators.

Keywords: natural language processing, combinatory categorial grammar, annotation tool

1. Pendahuluan

Latar Belakang

CCGweb[1] merupakan alat anotasi *open source* berbasis web pertama yang dikembangkan khusus untuk memberikan anotasi CCG[3]. Fitur yang ditawarkan CCGweb cukup beragam mulai dari *dynamic annotation*, WYSIWYG (*what you see is what you get*), *lexical category constraint*, *span constraint*, *issue reporting* via layanan *web service* eksternal, hingga *adjudication support*. Selain itu, CCGweb dibangun untuk membangun *dataset* CCG *multilingual* yang artinya terdapat lebih dari satu bahasa untuk satu kalimat yang sama. Untuk dapat menggunakan CCGweb, pengguna harus melakukan instalasi manual secara mandiri dan dapat dilakukan baik di komputer pribadinya ataupun di layanan *hosting* maupun *cloud*. Adapun proyek anotasi yang dapat dikerjakan dalam satu waktu adalah satu buah proyek yang mana dapat memiliki satu bahasa atau lebih. Demikian itu, pengguna tidak dapat mengerjakan lebih dari satu proyek dan harus menyelesaikan proyek sebelumnya agar dapat memulai proyek baru. Alternatif lainnya adalah melakukan instalasi kembali sehingga pengguna akan memiliki lebih dari satu aplikasi CCGweb.

Untuk dapat menggunakan CCGweb, pengguna diharapkan telah mempersiapkan beberapa *dependency* yang diperlukan yaitu EasyCCG[2] (termasuk berkas modelnya), Elephant *tokenizer*[3] (termasuk

[1]`https://github.com/texttheater/ccgweb`
[2]`https://github.com/ParallelMeaningBank/easyccg`
[3]`https://github.com/ParallelMeaningBank/elephant`