

Tugas Pekan ke-6 *Constituency Parsing*

Batas pengumpulan: Jumat 23 Oktober 2020, pukul 10.59 pagi, melalui LMS

Deskripsi

Lakukan eksplorasi pendefinisian *Grammar* dengan CFG, mengurai kalimat dengan *Grammar* yang sudah didefinisikan, dan membangkitkan *Grammar* dari *treebank constituency* Bahasa Indonesia.

Bagian I

1. Buatlah definisi sebuah *grammar* CFG sederhana untuk teks Bahasa Indonesia, yang mengandung ambiguitas.
2. Beri contoh 3 kalimat, lakukan *parsing*/penguraian berdasarkan *grammar* yang didefinisikan pada poin 1. Sertakan keterangan pohon hasil dari proses penguraian, beri keterangan jika terdapat ambiguitas.
3. Periksa apakah *grammar* yang Anda definisikan memenuhi syarat CNF.
4. Jika belum memenuhi syarat CNF, ubahlah menjadi CNF!

Bagian II

1. Lakukan induksi *grammar* PCFG dari *treebank constituency* Bahasa Indonesia <https://github.com/ialfina/kethu>
2. Berdasarkan 2 contoh kalimat:
 - a. Satu kalimat diambil dari daftar kalimat yang ada di *constituency Treebank* Bahasa Indonesia (boleh berupa potongan).
 - b. Satu kalimat diambil dari kalimat di luar contoh yang ada di *constituency Treebank* Bahasa Indonesia

Lakukan *parsing* dengan metode yang dapat menghasilkan semua kemungkinan pohon *parsing* (misal: BottomUp parser), dan yang menghasilkan hanya 1 pohon *parsing* (misal: Viterbi parser)

File yang harus dikumpulkan:

1. Program dan kelengkapannya: 1 file kode program python (.py) + petunjuk menjalankan program (.txt). (contoh kalimat dapat ditulis langsung/*hardcode*) di kode program
2. Laporan: 1 file pdf, maksimum panjang laporan adalah 2 halaman.

Penilaian: 20% source code + 80% laporan

Detail penilaian:

a. Program:

- kebenaran implementasi bagian I (*parsing* dan pemeriksaan apakah *Grammar* CNF) (10 poin)
- kebenaran implementasi bagian *parsing* bagian II (10 poin)

b. Laporan:

- kelengkapan eksperimen dan hasil eksperimen bagian I dan II (20 poin)
- analisis hasil parsing bagian I (10 poin)
- pemeriksaan CNF grammar bagian I (10 poin)
- konversi grammar non-CNF ke CNF bagian I (10 poin)
- analisis hasil *parsing* kalimat uji yang diambil dari *constituency Treebank* (15 poin)
- analisis hasil *parsing* kalimat uji **bukan** dari *constituency Treebank* (15 poin)

Jika ada pertanyaan, silakan disampaikan melalui *channel* pekan_6_tugas_constituency_parsing di slack.