

Tugas Besar Teori Bahasa dan Automata

Semester Genap 2021/2022

Membuat Lexical Analyzer dan Parser Sederhana untuk Teks Bahasa Alami

Deskripsi:

Pada Tugas Besar ini, mahasiswa (secara berkelompok; 2-3 orang per kelompok) diminta untuk mendefinisikan sebuah Context Free Grammar (CFG) yang merepresentasikan aturan bahasa sederhana untuk bahasa manusia. Kemudian berdasarkan CFG yang sudah didefinisikan tersebut, mahasiswa diminta untuk membuat sebuah **program sederhana berbasis web** (php/ java/ python/ .net/ dll.) untuk lexical analyzer (mengidentifikasi apakah sebuah lexical/token/kata valid sesuai simbol terminal yang didefinisikan) dan parser (apakah susunan token/kata sudah memenuhi aturan pada Grammar).

Berikut adalah contoh Grammar yang didefinisikan, yang merepresentasikan kalimat bahasa Indonesia yang valid secara sintaks.

Grammar:

```
<S> ::= <noun> <verb> <noun>  
<noun> ::= adik | kakak | buku | sepatu | bakso | tahu | topi  
<verb> ::= membaca | makan | memakai
```

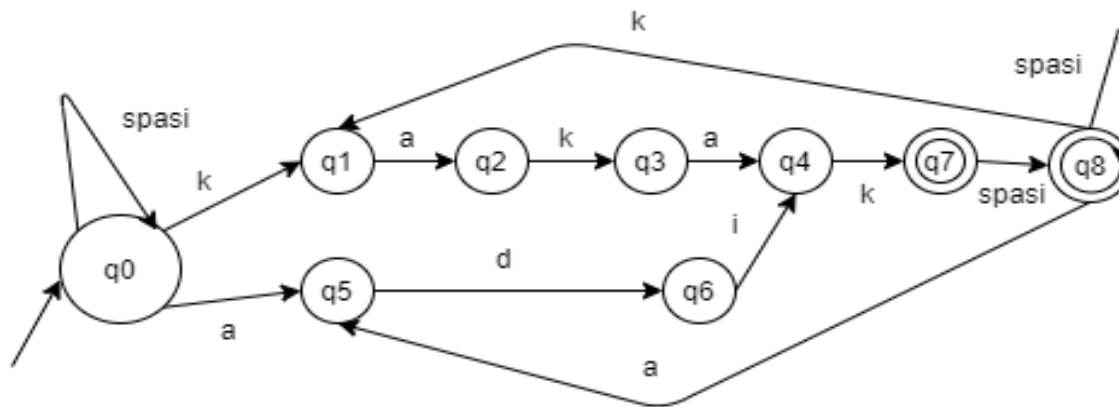
Simbol non-terminal: S (starting symbol), noun, verb

Simbol terminal: adik, kakak, buku, sepatu, bakso, tahu, membaca, makan, memakai

Catatan: notasi Grammar dituliskan menggunakan notasi BNF (Backus-Naur Form)

Kemudian berdasarkan Grammar tersebut, didefinisikan Finite Automata untuk mengenali lexical serta parse-table LL(1) sebagai berikut:

- Finite automata (**Perhatian**: ini contoh, hanya untuk 2 kata, nanti kalian harus buat lengkap untuk semua simbol terminal/kata!)



b. Parse table LL(1)

	kakak	adik	bakso	tahu	membaca	makan	memakai	buku	sepatu	topi	EOS
S	NN VB NN	NN VB NN	NN VB NN	NN VB NN	error	error	error	NN VB NN	NN VB NN	NN VB NN	error
NN	kakak	adik	bakso	tahu	error	error	error	buku	sepatu	topi	error
VB	error	error	error	error	membaca	makan	memakai	error	error	error	error

Berdasarkan Grammar dan parse-table, dibuat program lexical analyzer dan parser. Program akan diberikan sejumlah kalimat uji untuk diidentifikasi apakah lexical pada kalimat uji tersebut valid dan apakah kalimat tersebut sintaksnya sesuai dengan Grammar.

Tugas yang harus dikerjakan mahasiswa per kelompok (2-3 mahasiswa per kelompok):

1. Buatlah Context Free Grammar sederhana yang merepresentasikan aturan/sintaks kalimat dalam sebuah bahasa manusia tertentu, SELAIN Bahasa Indonesia, contoh: Bahasa Jawa, Bahasa Sunda, Bahasa Bali, Bahasa Italia, Bahasa Spanyol, dll. **Tipe bahasa yang disarankan adalah bahasa dengan struktur S-V-O (subject-verb-object), daftarnya antara lain dapat dilihat di: <https://en.wikipedia.org/wiki/Subject-verb-object> , supaya karakteristiknya mirip dengan yang diberikan di contoh. Masing-masing kelompok dalam satu kelas harus membuat CFG bahasa yang BERBEDA.** Jumlah simbol non-terminal minimal 3, dan harus mengandung: starting symbol, dan POSTag (kelas kata) utama VB dan NN. Jumlah simbol terminal (kata) minimal 10.
2. Buatlah rancangan Finite Automata untuk lexical analyzer kata/token/symbol terminal sesuai dengan yang didefinisikan pada Grammar.

3. Buatlah rancangan parse-table LL (1) sesuai Grammar yang sudah didefinisikan.
4. Buatlah program lexical analyzer dan parser sederhana berbasis web sesuai rancangan Finite Automata dan parse table yang sudah dibuat.
5. Lakukan pengujian pada program lexical analyzer dan parser yang sudah dibuat, dengan ketentuan sebagai berikut:
 - a. Kata uji untuk lexical analyzer berupa 3 kata yang valid dan 3 kata yang tidak valid berdasarkan daftar simbol terminal.
 - b. Kalimat uji untuk parser dengan panjang antara 2-5 kata, terdiri dari 3 kalimat yang sesuai dengan Grammar (diterima) dan 3 kalimat yang tidak sesuai dengan Grammar (tidak diterima).

Deliverables:

1. Laporan, yang memuat deskripsi penjelasan pengerjaan tugas poin **1,2,3,4, dan 5** di atas. [POIN maksimum: 60]
 - a. Deskripsi CFG dan rancangan Finite Automata dikumpulkan maksimal **Rabu, 8 Juni 2022**.
 - b. Laporan lengkap dikumpulkan maksimal **Rabu, 15 Juni 2022**.
2. Program lexical analyzer dan parser yang siap di-run dan dilengkapi penjelasan bagaimana cara menjalankannya. [poin maksimum: 40]
 - a. Program lexical analyzer dan petunjuk menjalankan dikumpulkan maksimal **Rabu, 8 Juni 2022**.
 - b. Program Parser dan petunjuk menjalankan dikumpulkan maksimal **Rabu, 15 Juni 2022**.

Bonus Penilaian:

1. Program dapat menerima input langsung dari pengguna (tidak di-hardcode di dalam badan program): 10 poin
2. Penggabungan program lexical analyzer dan parser (hanya sekali pemanggilan untuk sebuah kalimat uji, akan diproses di lexical analyzer dulu dan kemudian bila valid akan diproses oleh parser): 10 poin.

Link Contoh Program dalam Bahasa Python (Desktop) : <https://www.youtube.com/playlist?list=PLsQUw5vHd9DQHSL4WdsE2HK6iK1nI5ocM>