Compiler Construction

Dr. Michael Petter, Raphaela Palenta SS 2019

Exercise Sheet 3

Assignment 3.1 Derivation orders

For the example grammar with start symbol E:

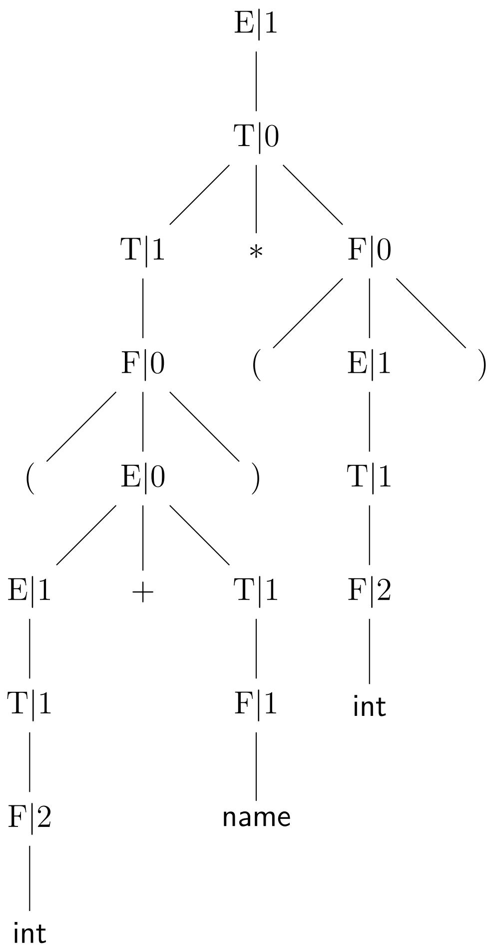
$$\begin{aligned}
E &\to E + T\ ^{0} \mid T\ ^{1}\\
T &\to T * F\ ^{0} \mid F\ ^{1}\\
F &\to (E)\ ^{0} \mid \mathsf{name}\ ^{1} \mid \mathsf{int}\ ^{2}
\end{aligned}$$

Draw derivation trees and give yields for the following

- leftmost derivation:
  (E,1) (T,0) (T,1) (F,0) (E,0) (E,1) (T,1) (F,2) (T,1) (F,1) (F,0) (E,1) (T,1) (F,2)
- reverse rightmost derivation:
  (F,2) (T,1) (F,1) (T,0) (E,1) (F,1) (T,1) (F,1) (T,0) (E,0)

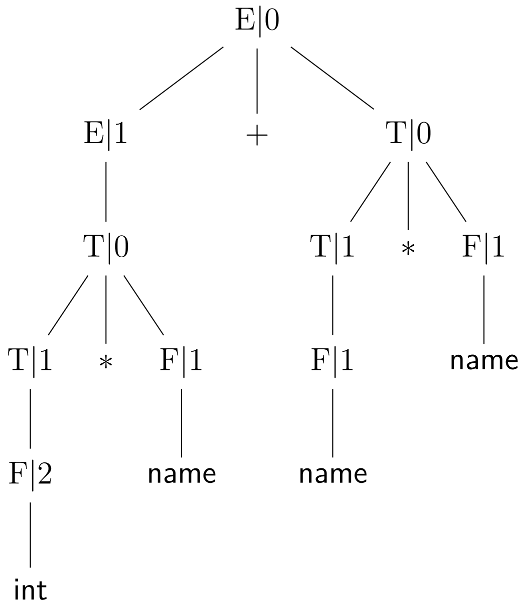
Suggested Solution 3.1

Let t equal the following tree



then $\mathsf{yield}(t) = (\mathsf{int} + \mathsf{name}) * (\mathsf{int})$

Let t' equal the following tree



then $\mathsf{yield}(t') = \mathsf{int} * \mathsf{name} + \mathsf{name} * \mathsf{name}$

Assignment 3.2 Uniqueness

For each grammar G below with start symbol A.

- What is the language of the grammar (use the set notation)?
- Is the language of the grammar a regular language?
- Is the grammar unique?
- If the grammar is not unique (ambiguous) give two derivation trees with the same yield as a counterexample.
- If the grammar is not unique (ambiguous) give an equivalent unique grammar.

1.
$$A \to AA\ ^0 \mid a\ ^1$$

2.
$$\begin{aligned} A &\to aA\ ^0 \mid a\ ^1 \mid B\ ^2 \\ B &\to bB\ ^0 \mid b\ ^1 \mid A\ ^2 \end{aligned}$$

3.
$$\begin{aligned} A &\to aB\ ^0 \mid b\ ^1 \\ B &\to aA\ ^0 \end{aligned}$$

4.
$$\begin{aligned} A &\to BA\ ^0 \mid \epsilon\ ^1 \\ B &\to Bb\ ^0 \mid aBb\ ^1 \mid ab\ ^2 \end{aligned}$$

Suggested Solution 3.2

1. - $\mathcal{L}(G) = \{a^n \mid n \geq 1\}$
   - $\mathcal{L}(G)$ is regular, i.e., $\mathcal{L}(G) = [\![a^+]\!]$
   - G is not unique since $\mathsf{yield}(A0(A1(a), A0(A1(a), A1(a)))) = \mathsf{yield}(A0(A0(A1(a), A1(a)), A1(a))) = aaa$
   - $A \to Aa \mid a$

2.
- $\mathcal{L}(G) = \{w_1 \dots w_n \mid w_i \in \{a, b\}, n \geq 1\}$
- $\mathcal{L}(G)$ is regular, i.e., $\mathcal{L}(G) = [\![(a \mid b)^+]\!]$
- G is not unique as $\mathsf{yield}(A2(B2(A1(a)))) = \mathsf{yield}(A1(a)) = a$
- $A \rightarrow aA \mid bA \mid a \mid b$

3.
- $\mathcal{L}(G) = \{a^{2n}b \mid n \geq 0\}$
- $\mathcal{L}(G)$ is regular, i.e., $\mathcal{L}(G) = [\![(aa)^*b]\!]$
- unique as the grammar is reduced and right linear

4.
- $\mathcal{L}(G) = \{w_1 \dots w_n \mid w_i \in L, n \geq 0\}$ with $L = \{a^n b^m \mid 1 \leq n \leq m\}$
- $\mathcal{L}(G)$ is not regular. We apply the pumping lemma for regular languages. Let $x = a^n b^m = uvw$ with $u = a^{n_1}$, $u = a^{n_1}$, $v = a^{n_2}$ and $w = a^{n_3}b^m$, $n_1 + n_2 \leq n$, $n_2 \geq 1$, $n_1 + n_2 + n_3 = n$. Then $uv^{m+1}w$ would be a word in L. But $uv^{m+1}w = a^{n_1}a^{n_2(m+1)}a^{n_3}b^m$ with $n_1 + n_2(m+1) + n_3 > m$, a contradiction.
- G is not unique since
  $\mathsf{yield}(A0(B1(a, B0(B2(a, b), b), b), A1(\epsilon))) =$
  $\mathsf{yield}(A0(B0(B1(a, B2(a, b), b), b), A1(\epsilon))) = aabbb$
- $$\begin{aligned} A &\rightarrow BA \mid \epsilon \\ B &\rightarrow Bb \mid C \\ C &\rightarrow aCb \mid ab \end{aligned}$$

Assignment 3.3 Berry-Sethi is still alive!

Last week we implemented the sophisticated Berry-Sethi approach in order to derive an NFA from a regular expression. This week we implement the next step in order to derive a (partial) DFA from the NFA for a given word and check if this is accepted by the NFA/DFA or not.

Your task is to implement a method `String getPartialDFA(String in)` which returns a (partial) DFA from the NFA for a given word `in`. If you call the method twice or several times in row, the previously computed (partial) DFA should be reused and extended if necessary. Again we represent our resulting automata by the Graphviz DOT language. The method should print on the console if the word is accepted or not.