CS143 Spring 2022 – Written Assignment 2

Monday, April 25, 2022 11:59 PM PDT

This assignment covers context free grammars and parsing. You may discuss this assignment with other students and work on the problems together. However, your write-up should be your own individual work, and you should indicate in your submission who you worked with, if applicable. Assignments can be submitted electronically through Gradescope as a PDF by 11:59 PM PDT. Please review the the course policies for more information: `https://web.stanford.edu/class/cs143/policies/`. A LaTeX template for writing your solutions is available on the course website. If you need to draw parse trees in LaTeX, you may use the `forest` package: `https://ctan.org/pkg/forest`.

1. Give a context-free grammar (CFG) for each of the following languages. Any grammar is acceptable—including ambiguous grammars—as long as it has the correct language. The start symbol should be S.

   (a) The set of all strings over the alphabet $\{1, 2, *\}$ representing valid products of integers where the expression evaluates to some even value.
   Example Strings in the Language:

   112   2*121   221*1*1122

   Strings not in the Language:

   ε 11*121   12**22*112

   (b) The set of all strings over the alphabet $\{x, (,), ;\}$ representing nested tuples of x's where each tuple has an even length.
   Example Strings in the Language:

   $()$ $(x; ())$ $((); x; ((); x); x)$

   Strings not in the Language:

   ε x $((); x; x)$ $(x; (); (x; (); x); x)$

   (c) The set of all strings over the alphabet $\{0, 1\}$ where the number of 1's is exactly one more than the number of 0's.
   Example Strings in the Language:

   1   101   001110101

   Strings not in the Language:

   ε 001   01100101

   (d) The set of all strings over the alphabet $\{0, 1\}$ in the language $L : \{0^i 1^j 0^k \mid j \neq i + k\}$.
   Example Strings in the Language:

   00   110000   000111110

   Strings not in the Language:

   ε 0011   01111000

2. Consider the following grammar for if-then-else expressions that involve a variable x:

$$
\begin{aligned}
E &\to \text{if } x \text{ then } E \mid M \\
M &\to \text{if } x \text{ then } M \text{ else } E \mid x
\end{aligned}
$$

Is this grammar ambiguous or not? If yes, give an example of an expression with two different parse trees and draw the two parse trees. If not, explain why that is the case.

3. (a) Eliminate left recursion from the following grammar:

$$
\begin{aligned}
S &\to S(T) \mid Sa \mid [T] \mid Tb \\
T &\to T(S) \mid Tc \mid d
\end{aligned}
$$

(b) Left factor the following grammar:

$$
\begin{aligned}
S &\to (T+T) \mid (T) \\
T &\to U * T \mid U*? \mid [U] \\
U &\to U0 \mid U1 \mid \varepsilon
\end{aligned}
$$

4. Consider the following CFG, where the set of terminals is $\{a, b, c, d, \#, ?\}$:

$$
\begin{aligned}
S &\to \#UT \mid T? \\
T &\to aS \mid bUc \mid \varepsilon \\
U &\to aSc \mid bTd
\end{aligned}
$$

(a) Construct the FIRST sets for each of the nonterminals.

(b) Construct the FOLLOW sets for each of the nonterminals.

(c) Construct the LL(1) parsing table for the grammar.

(d) Show the sequence of stack, input and action configurations that occur during an LL(1) parse of the string "$\#a?ca?$". At the beginning of the parse, the stack should contain a single S.

5. Consider the following grammar G over the alphabet $\{a, b, c\}$:

$$
\begin{aligned}
S' &\to S \\
S &\to Aa \\
S &\to Bb \\
A &\to Ac \\
A &\to \varepsilon \\
B &\to Bc \\
B &\to \varepsilon
\end{aligned}
$$

You want to implement G using an SLR(1) parser. Note that we have already added the $S' \to S$ production for you.

(a) Construct the first state of the LR(0) machine, compute the FOLLOW sets of A and B, and point out the conflicts that prevent the grammar from being SLR(1).

(b) Show modifications to production 4 ($A \to Ac$) and production 6 ($B \to Bc$) to make the grammar SLR(1) while having the same language as the original grammar G. Explain the intuition behind this result.