CSE340 Fall 2020 - Homework 1

All submissions should be typed. Exception can only be made for drawing parse trees, which can be hand drawn and scanned in the submitted document.

When you submit your solution on Gradescope, you should indicate for each problem the page on which the solution is. Remember that no late submissions are accepted for homework assignments.

Problem 1. Consider the list of tokens

T1 = { "abc", abcd1e" }

T2 = { "abd" }

ID = Set of strings that consist of a letter or underscore that is followed zero or more letters, underscore or digits.

NUM = Set of strings that consist of a non-zero digit that is followed by 1 or more digits or the string "0".

Consider the input

abcd22abc 123 00abc abd

and the following sequence of calls:

t1 = lexer.GetToken();
t2 = lexer.GetToken();
t3 = lexer.peek(1);
t4 = lexer.peek(3);
t5 = lexer.peek(5);
t6 = lexer.peek(7);
lexer.UngetToken(2);
t7 = lexer.GetToken();
t8 = lexer.GetToken();

Assume that space is a separator, but is otherwise ignored.

1. What are the values of t1, t2, t3, t4, t5 and t6?

2. What are the values of t7 and t8?

Problem 2. Consider the grammar

$S \rightarrow A\,B\,C$
$A \rightarrow a\,A\,b \mid \varepsilon$
$B \rightarrow b\,B \mid b$
$C \rightarrow b\,C\,a \mid b$

1. What is the start symbol? Explain briefly!

2. What are the non-terminals? Explain briefly!

3. What are the terminals? Explain briefly!

4. Give a parse tree for the input:

   a a b b b b b b a a

5. In the parse tree of

   a a b b b b b b a a

   the root node is labeled S and its children are labeled A, B and C from left to right. Which parts of the input correspond to the children of S in the parse tree? You answer should have the following format:

   The part of input that corresponds to A is …
   The part of input that corresponds to B is …
   The part of input that corresponds to C is …

Problem 3. Consider the grammar

$A \rightarrow X\,Y \mid Z\,X$
$X \rightarrow a\,B \mid Y$
$B \rightarrow b\,B \mid X \mid \varepsilon$
$Y \rightarrow a\,Y\,b \mid B$
$Z \rightarrow a\,Z\,b \mid X$

where A, B, X, Y and Z are non-terminal, A is the start symbol and a and b are tokens. Remember that ε represent the empty string. $Y \rightarrow \varepsilon$ means that Y does not have to match any tokens.

1. Give a parse tree for the sequence of tokens:

   a b a a b a

2. Give a another (different) parse tree for the sequence of tokens:

   a b a a b a

3. Is this grammar ambiguous? Why?