Middle East Technical University Department of Computer Engineering

CENG 280

Formal Languages and Abstract Machines

Spring 2016-2017

Take Home Exam 2

Due date: April 29, 2017, Sat, 23:55

Objectives

To familiarize with context-free languages, the class of languages whose strings can be generated by context-free grammars which are tools based on more elaborated understanding of the structure of member strings.

Specifications

This take-home exam comprises of seven questions, and only the first six questions will be graded. You do not need to submit your solution of the seventh question, which includes the topics to be covered just before the announced deadline of the take-home exam, yet you must be prepared to solve similar questions for the second midterm.

Questions are included in subsequent sections. Concerning submission rules and general regulations governing your side, you may refer to the last two sections of this document.

In the first question, you must define a DFA to recognize the given language and then convert this into an equivalent DFA with the minimal number of states by making use of Myhill-Nerode Theorem, which you will in turn utilize to define equivalence classes partitioning the set of all strings via the relation imposed by the minimal DFA.

In the second question, you must use Myhill-Nerode Theorem to deduce that the given language is not regular, then you must define a PDA to recognize it as a CFL on which you should trace the given strings.

For the third question, you should identify the language generated by the given CFG, which must further be inspected for being ambiguous or not. In case of detected ambiguity, you should propose an unambiguous CFG for the same language and comment on why you think it is unambiguous.

Fourth question asks you to define a CFG for the specified language and you should give example CFL's for the cases indicated in the fifth question. The sixth question invites you to deduce whether given languages are CFL's or not, involving an application of pumping lemma for CFL's for the latter case.

Designing your solutions to the tasks, explicitly state any assumptions you make and pay particular attention to the notation you use. Try to minimize informal subjective comments in your work.

Question 1 (10 pts)

You are given the following regular language:

$L = \{\ xy : |x| = 2|y|,\ x, y \in \{a, b\}^*,\ |y| \geq 1$ and the second last letter of xy is $b.\ \}$

a. Construct a DFA to recognize L and minimize the number of states it has.

b. Define the equivalence classes partitioning the set $\{a, b\}^*$ imposed by the minimal DFA as regular expressions.

Question 2 (20 pts)

You are given the following language:

$L = \{\ xy : |x| = |y|,\ x, y \in \{0, 1\}^*$, and x contains a substring 01. $\}$

a. Using Myhill-Nerode theorem, prove that L is not regular.

b. Prove that L is a CFL by constructing a PDA M such that $L(M) = L$.

c. Show whether the following strings are in L or not by tracing the sequence of moves of PDA M on:

i. $w_1 = 10100001$

ii. $w_2 = 1101$

Question 3 (20 pts)

Assume that $G = (V, \Sigma, R, S)$ is a CFG with $V = \{S, T, a, b\}$, $\Sigma = \{a, b\}$, and
$R = \{\ S \rightarrow aaTb,\ T \rightarrow aTb,\ T \rightarrow S,\ T \rightarrow e\ \}$.

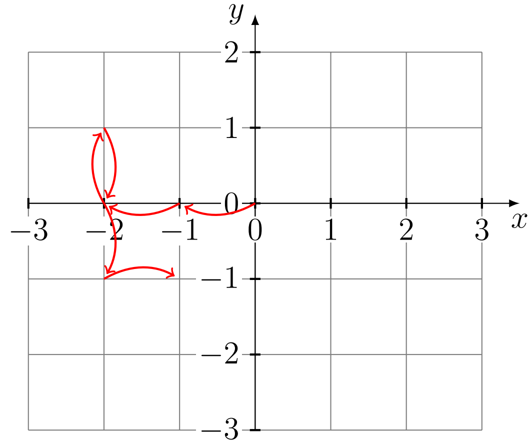
a. Let $L(G) = \{\ w \in \{a, b\}^* : w$ has property $P\ \}$. Identify what P is.

b. Write down all rightmost derivations of $w = aaaabbb$ by G and draw associated parse trees.

c. Is G ambiguous? If so, suggest an unambiguous CFG G_u s.t. $L(G_u) = L(G)$ and discuss how G_u can be proven to be unambiguous. Otherwise discuss why the given CFG G is unambiguous. If you think $L(G)$ is inherently ambiguous, state reasons for your opinion.

Question 4 (15 pts)

Consider the alphabet $\Sigma = \{u, d, l, r\}$, where the letters represent going one unit step to up, down, left and right directions in $\mathbb{R}^2$, respectively. Set the starting point to origin for convenience. Notice that each string in Σ^* corresponds to a path in $\mathbb{R}^2$. As an example, the path that corresponds to the string *lluddr* is given in the following.



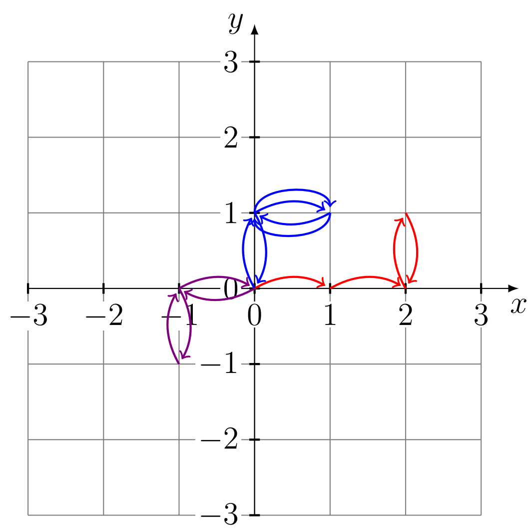
The path which corresponds to *lluddr*

Define "unit line segment" as an undirected line segment whose length is 1, i.e. the line segment between (x_0, y_0) and (x_0, y_0+1) or between (x_0, y_0) and (x_0+1, y_0), where $x_0, y_0 \in \mathbb{R}$. You are expected to define a context-free grammar for the language L whose alphabet is Σ, and which consists of all strings that satisfy the following conditions:

i. In the corresponding path, each unit line segment must be passed an even number of times.

ii. For each unit line segment in the corresponding path, each odd numbered pass must be in up or right direction, and each even numbered pass must be in down or left direction.

For example, the string $rrud \notin L$, because the unit line segments between (0,0)-(1,0) and (1,0)-(2,0) don't satisfy the condition (i), and the string $ldur \notin L$, because the unit line segments between (0,0)-(-1,0) and (-1,0)-(-1,-1) don't satisfy the condition (ii). The empty string, and the strings *urlrld* and *udrl* are in L, because they satisfy both conditions.



The paths which correspond to $rrud \notin L$ (in red), $ldur \notin L$ (in violet) and $urlrld \in L$ (in blue)

Question 5 (9 pts)

Formally define example CFL's for the following questions.

a. Find a CFL that is not a regular language, whose complementation is also a CFL.

b. Find two non-regular CFL's whose intersection would yield another non-regular CFL.

c. Find two CFL's whose intersection would yield a language that is not context-free.

Question 6 (26 pts)

Decide whether the following languages are context-free or not. If you think that a language is a CFL, provide a CFG generating strings of the language. Otherwise, use pumping lemma for CFL's to prove that the language is not context-free. You can refer to closure properties of CFL's in your proofs. Assume that $n_x(w)$ is a function from the set of strings to the integers, returning the number of occurrences of the character x within the string w.

a. $L_1 = \{\ 0^m 1^n : m = 3^k,\ n = k^2,\ k \geq 0\ \}$

b. $L_2 = \{\ w \in \{a, b, c\}^* : n_a(w) \leq n_b(w) \text{ or } n_a(w) > n_c(w)\ \}$

c. $L_3 = \{\ w \in \{a, b, c\}^* : 3n_a(w) = 5n_b(w) = 2n_c(w)\ \}$

Question 7 (Not to be graded)

Assume that $G = (V, \Sigma, R, S)$ is a CFG with $V = \{S, A, B, T, a, b, \$\}$, $\Sigma = \{a, b, \$\}$, and $R = \{\ S \rightarrow AB,\ A \rightarrow aAa \,|\, bAb \,|\, T,\ B \rightarrow aB \,|\, bB \,|\, e,\ T \rightarrow \$\$B\ \}$.

a. Convert G into some equivalent CFG in CNF.

b. Using CYK algorithm, decide whether the following strings belong to $L(G)$.

i. $w_1 = ab\$\$aaab$

ii. $w_2 = ba\$\$aaba$

Submission

- **Late Submission:** You have 3 days in total for late submission of all homeworks. All homeworks will be graded as normal during this period. No further late submissions are accepted.
- You should submit your THE2 as a .tex file. Please use the template provided on COW with appropriate modifications.
- Soft-copies should be uploaded strictly by the deadline.

Regulations

1. **Cheating: We have zero tolerance policy for cheating**. People involved in cheating will be punished according to the university regulations.

2. **Newsgroup:** You must follow the newsgroup (news.ceng.metu.edu.tr) for discussions and possible updates on a daily basis.