Maggie Johnson
CS109

Handout #26

due: 3/10, 100 points

Problem Set #7

We can just skim the surface of formal languages in CS109. It's a very important area of study in both theoretical computer science, and in programming language design and compiler implementation. Our goal in this problem set is to make a start at developing skills in using regular expressions and context-free grammars to represent specific patterns in languages. Be sure to keep the notations of the two separate, e.g., there are no *'s or +'s in grammar definitions (unless these symbols are in the language being expressed). Also, be sure that your finite automata are *strictly* deterministic since that is the only form of which we have learned.

There is no need to show your work on this problem set – just give the regular expression, finite automata or grammar. If we ask you to prove something, be sure to be complete and precise (as usual).

Regular Expressions and Finite Automata

1) Express each of the following (over the alphabet {0, 1}) using regular expressions:
   a. The set of strings of one or more 0s followed by a 1. (3 points)
   b. The set of strings of two or more symbols followed by three or more 0s. (3 points)
   c. The set of strings with either no 1 preceding a 0 or no 0 preceding a 1. (4 points)
   d. The set of strings containing one or more 1s followed by an even number of 0s. (4 points)
   e. The set of strings not containing a "11". (5 points)
   f. The set of all strings except the string "11". (5 points)

2) Describe (precisely) in English the following languages over the alphabet {0, 1}:
   a. ((01 + 10)* + (11+00)*)* (3 points)
   b. (0+1)* (0+10) (3 points)

3) (10 points) Consider the set of strings representing fixed-point positive numbers, i.e. strings like 34.21 or 234 or 0.006. Each allowable string is either (i) a non-empty string of digits or (ii) a non-empty string of digits followed by a decimal point and by another non-empty string of digits. For example, "234" and "234.0" are legal but ".234" and "234." are not. Furthermore, numbers between 0 and 1 should have a single leading 0 as in "0.005" but no more ("00.04" is illegal) and other numbers should not have leading zeroes ("023" is illegal). Write a deterministic finite automata that accepts this set of strings.

4) (10 points) Define a finite automata for the following language over the alphabet {0, 1}: the set of all strings with an equal number of 0's and 1's, such that each prefix has at most one more 0 than 1's, and at most one more 1 than 0's.

Context-Free Grammars

5) (5 points each) Define a context-free grammar for the following over the alphabet {a,b}:

a) the set of strings of the form a^{2n} b^{3n} where n is a non-negative integer.

b) the set of strings generated by a*b*a*

c) the set of strings made up of an a followed by an odd number of b's

d) the set of strings containing an even number of a's

e) A palindrome is a word that reads the same forwards and backwards. For example, "mom" or "sees" are palindromes. Define a context-free grammar that generates all palindromes over the alphabet {a, b}

6) (10 points) Define a context-free grammar over the alphabet {b, c} where every string has twice as many b's as c's.

7) Let G_1 and G_2 be context-free grammars, generating the languages $L(G_1)$ and $L(G_2)$ respectively. Prove that there is a context-free grammar generating each of the following sets. (5 points each)

a) $L(G_1) \cup L(G_2)$

b) $L(G_1)\, L(G_2)$

c) $L(G_1)^*$