Midterm Examination
CS 4232: Theory of Computation

Thursday 29 October 2015, Duration 40 Minutes

Student Number: ________________

Rules: This test carries 20 marks and consists of 4 questions. Each questions carries 5 marks; full marks for a correct solution; a partial solution can give a partial credit. Use the backside of the page if the space for a question is insufficient.

Question 1 [5 marks]

Let $L = \{5^n \cdot 6^m \cdot 7^k : n \geq m \text{ and } m \geq k\}$ and let $h : \{0,1,2,3\}^* \to \{5,6,7\}^*$ be a homomorphism given by $h(0) = 5$, $h(1) = 5$, $h(2) = 55$, $h(3) = 6$. Consider the set

$$H = h^{-1}(L) \cap (\{1\}^* \cdot \{2\}^* \cdot \{3\}^*).$$

Determine the best possible position of H in the Chomsky hierarchy (regular, context-free, context-senstive, recursively enumerable) and construct the corresponding grammar.

Solution. The set H consists of all words $1^h 2^i 3^j$ such that $h + 2i \geq j$, as a 1 is mapped to one 5 and a 2 to two 5s while a 3 is mapped to a 6. The choice is therefore context-free and the grammar would look as follows: The non-terminals are S, T, the terminals are $0, 1, 2, 3$, the start symbol is S and the rules are

- $S \to 1S3|1S|T$,
- $T \to 2T33|2T3|2T|\varepsilon$.

Note that $h(2) = 55$ which can therefore balance out two 3, as $h(3) = 6$.

It is known that the language $L = \{w \in \{0,1,2\}^* : w$ contains strictly more 0 than 1$\}$ is deterministic context-free. What about the languages $H = L \cup \{0^n 1^n 2^n : n \in \mathbb{N}\}$ and $K = (L \cap (\{0\}^* \cdot \{1\}^* \cdot \{2\}^*)) \cup \{2,3\}^*$? Prove your answer; theorems from the lecture can be used.

Solution. The language $H = L \cup \{0^n 1^n 2^n : n \in \mathbb{N}\}$ is not context-free and therefore also not deterministic context-free. Assume by way of contradiction that H satisfies the context-free pumping lemma with constant n. Let $u = 0^n 1^n 2^n$ and let $vwxyz$ be the word u splitted according to the pumping lemma. If w, y contains the same amount of $0, 1$ then $vwwxyyz$ must also be of the form $0^m 1^m 2^m$ for some $m > n$ and this is impossible as the pump cannot contain any 2 by $|wxy| \leq n$. If w, y contain more 0 than 1 then vxz contains less 0 than 1 and is not in H so that again the pumping lemma is not satisfied. If w, y contain less 0 than 1 then $vwwxyyz$ contains less 0 than 1 and is not in H and again the pumping lemma is not satisfied.

The language $L \cap (\{0\}^* \cdot \{1\}^* \cdot \{2\}^*)$ is the intersection of a deterministic context-free language and a regular language and thus deterministic context-free. Now K is the union of that language and a regular language and also determinitistic context-free.

Question 3 [5 marks]

Write a register machine program which computes the function F given by $x \mapsto x^2+2x+5$; the register machine can use addition, comparison, subtraction, conditional jump and unconditional jump; note that the register machine has always natural numbers in the variables and that therefore $2-5$ is 0 and not -3. All macros used must be defined within this question by their own programs.

Solution. The following program computes the function. R_2 is a counter which goes from 0 to the input R_1 and R_3 holds each time in Line 4 the value $R_2 \cdot (R_1+2)+5$ which at the end will be $R_1^2+2R_1+5$ when the program gives the return value $\mathrm{F}(R_1)$ in Line 8.

Line 1: Function $\mathrm{F}(R_1)$;
Line 2: $R_2 = 0$;
Line 3: $R_3 = 5$;
Line 4: If $R_2 = R_1$ Then Goto Line 8;
Line 5: $R_2 = R_2 + 1$;
Line 6: $R_3 = R_3 + R_1 + 2$;
Line 7: Goto Line 4;
Line 8: Return(R_3).

Consider the grammar

$$(\{S, T, U\}, \{0, 1\}, \{S \to TT|UU, T \to UU, U \to 0|1\}, S)$$

and apply the algorithm of Cocke, Kasami and Younger to check whether the word 1001 is in the language L generated by the grammar. Furthermore, determine how many words L contains.

Solution. The table of the algorithm looks as follows for the word 1001:

$$\begin{array}{ccccccc} & & & E_{1,4} = \{S\} & & & \\ & & E_{1,3} = \emptyset & & E_{2,4} = \emptyset & & \\ & E_{1,2} = \{S,T\} & & E_{2,3} = \{S,T\} & & E_{3,4} = \{S,T\} & \\ E_{1,1} = \{U\} & & E_{2,2} = \{U\} & & E_{3,3} = \{U\} & & E_{4,4} = \{U\} \\ 0 & & 0 & & 1 & & 1 \end{array}$$

Thus the word is in the language. Indeed, one can see that U can generate every word in $\{0,1\}$. Furthermore, S and T can both generate UU and thus generate every word in $\{0,1\}^2$. As there is also the rule $S \to TT$, S can also generate every word in $\{0,1\}^4$ by deriving $S \Rightarrow TT \Rightarrow UUT \Rightarrow UUUU$ and then onward to any word of four bits by replacing each U according to the target. Thus the language contains all two-bit and all four-bit words which are in total 20 words.

END OF PAPER