INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
Mid-Spring Semester 2022-23

Date of Examination: 20/02/23 **Session(FN/AN)** AN **Duration** 2 hrs, **Marks** = **50**
Sub No: CS21004/CS21204 **Sub Name:** Formal Language & Automata Theory
Department/Centre/School : Computer Science and Engineering
Specific charts, graph paper,log book etc. required NO
Special Instructions (if any) ANSWER ALL questions. In case of reasonable doubt, make assumptions and state them upfront. Marks will be deducted for sketchy proofs and claims without proper reasoning. All parts of a single question should be done at the same place.

1. Consider the DFA in Figure 1. Show that the language of this DFA is equivalent to the regular expression (a + b(b + ab)*aa)*. [**4 marks**]

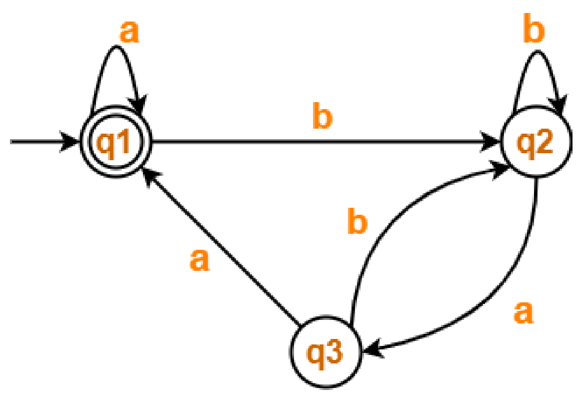


Figure 1: A sample DFA.

2. Assume l=letters, d=digits and _=standard underscore. Write a regular expression that admits only a valid variable name for C language. [**3 marks**]

3. Consider the alphabet $\sum = \{0, 1\}$ and the sets of strings generated by the corresponding non-terminals X_0, X_1, and X_2 of a regular grammar whose production rules are stated below.
$X_0 \to 1X_1$
$X_1 \to 0X_1|1X_2$
$X_2 \to 0X_1|\epsilon$
The start symbol is X_0. Construct the NFA corresponding to the above grammar. Hence, obtain the regular expression.
[**3+3 marks**]

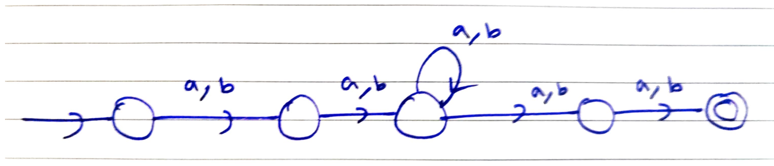
4. Consider the regular expression (a + a(b + aa)*b)*a(b + aa)*a. What is the minimum number of states in the NFA corresponding to this regular expression? Construct this NFA. What is the minimum number of states in the equivalent DFA. Draw this DFA. [**1+2+1+2 marks**]

5. Suppose that language A_1 has a context-free grammar $G_1 = (V_1, \sum, R_1, S_1)$, and language A_2 has a context-free grammar $G_2 = (V_2, \sum, R_2, S_2)$, where, for $i = 1, 2, \ldots, V_i$ is the set of variables, R_i is the set of rules/productions, and S_i is the start variable for CFG G_i. The CFGs have the same set of terminals $\sum$. Assume that $V_1 \cap V_2 = \phi$. Define another CFG $G_3 = (V_3, \sum, R_3, S_3)$ with $V_3 = V_1 \cup V_2 \cup \{S3\}$, where $S_3 \notin V_1 \cup V_2$, and $R_3 = R_1 \cup R_2 \cup \{S_3 \to S_1, S_3 \to S2\}$. Let the language generated by G_3 be A_3. Express A_3 in terms of A_1 and A_2. Justify formally with a proof. [Present step by step

proof showing the necessary and sufficient directions separately. Sketchy proofs will not be evaluated.]

[1+5 marks]

6. Prove that $L = \{vwv \mid v, w \in \{a,b\}^*, |v| = 2\}$ is regular. **[4]**



7. Use Pumping Lemma to prove $L = \{w \in \{0,1\} \mid www = uu$ for some u $\in \{0,1\}^*\}$ is not regular. **[5]**
Solution: Let N be the pumping length. Now the string $w = 10^N110^N1 \in L$ since $10^N110^N1 \cdot 10^N110^N1 \cdot 10^N110^N1 = 10^N110^N110^N1 \cdot 10^N110^N110^N1$. Now decomposing $w = xyz = 10^N110^N1$ with $|xy| \leq N$ in all possible ways will lead to the y segment comprising 0s only from the initial segment of 0s. Pumping this string will lead to a form $(10^k)^j0^{N-k}110^N1$ or $10^{N+i}110^N1$. In all such cases the symmetry of the string gets lost and it is not possible to divide 3 consecutive concatenations of the string to two identical concatenations.

8. (a) State the Myhill-Nerode Theorem.
   (b) Apply the Myhill-Nerode Theorem to prove that the language $\{a^nb^n \mid n \geq 0\}$ for $\Sigma = \{a,b\}$ is not regular. Do not use any other method.
   Solution: Check slides
   (c) Use the previous result to prove that $L = \{a^nb^l \mid n \neq l, n \geq 0, l \geq 0\}$ for $\Sigma = \{a,b\}$ is not regular.
   Solution: First prove that a^*b^* is regular. Observe that $\{a^nb^n \mid n \geq 0\} = a^*b^* \cap L^C$. As regular languages are closed under intersection and complementation, if L is regular then $\{a^nb^n \mid n \geq 0\}$ also becomes regular. Which is not true as shown in (b). Hence, ...

[2+4+2]

9. (a) Prove that for any right-linear grammar G there exists a strictly right-linear grammar H such that $L(G) = L(H)$. **[4]**
   Solution: In a right linear grammar, consider any rule which is either of type $A \to xB$ or $A \to x$, where $A, B \in V$, $x \in \Sigma^*$. Let $x = a_1 \cdots a_n$ where, $a_1, \cdots, a_n \in \Sigma$. Introduce new non-terminals $A_1, \cdots, A_n$ and new rules,
   $A \to a_1A_1$
   $A_1 \to a_2A_2$
   $\vdots$
   $A_n \to B$
   Thus the right linear rule $A \to a_1 \cdots a_nB$ can be replaced by the above strictly right linear rules. For the rule of type $A \to a_1 \cdots a_n$, use same method with the last introduced rule being $A_n \to \epsilon$. Doing similarly for all rules in G, one can derive H.
   (b) Find a regular grammar for $L = \{w \in \{a,b\}^*$ such that $|(n_a(w) - n_b(w))|$ is odd$\}$.
   (notations $n_a(w), n_b(w)$ represent the number of a, b in string w respectively and '$|x|$' represents magnitude of some signed integer x) **[4]**
   Solution:
   $S \to aA|bA; A \to aS|bS|\lambda$