CS21004 Formal Languages and Automata Theory, Spring 2011–12

Mid-Semester Test

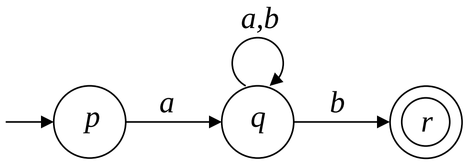
Maximum marks: 35 | Date: February 2012 | Duration: Two hours

---

Roll no: ________________ **Name:** ____

[*Write your answers in the question paper itself. Be brief and precise. Answer all questions.*]

1. Use the subset-construction procedure to convert the following NFA (over $\{a, b\}$) to an equivalent DFA. Mark the states of your DFA by subsets of $\{p, q, r\}$. Indicate which states of your DFA are accessible. No credit will be given if you design your DFA by any method other than the subset-construction procedure. **(5)**



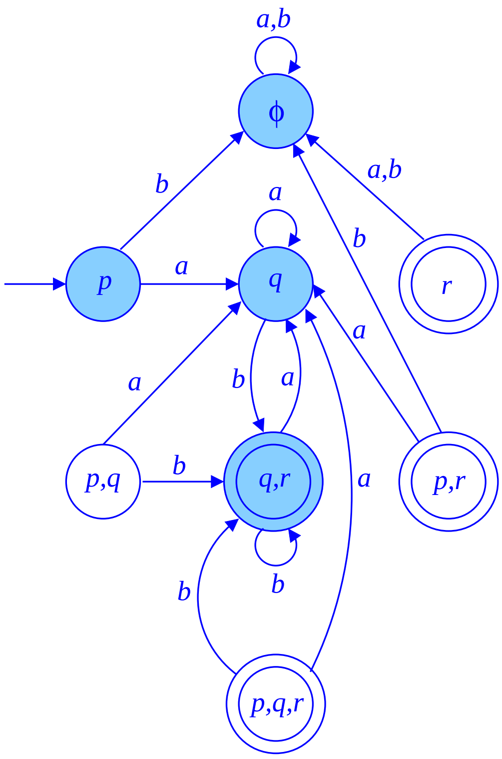
Solution We have

$$\Delta(p,a) = \{q\}, \quad \Delta(q,a) = \{q\}, \quad \Delta(r,a) = \emptyset,$$
$$\Delta(p,b) = \emptyset, \quad \Delta(q,b) = \{q,r\}, \quad \Delta(r,b) = \emptyset.$$

Therefore, we have

$$\Delta(\{p,q\},a) = \{q\}, \quad \Delta(\{q,r\},a) = \{q\}, \quad \Delta(\{p,r\},a) = \{q\}, \quad \Delta(\{p,q,r\},a) = \{q\},$$
$$\Delta(\{p,q\},b) = \{q,r\}, \quad \Delta(\{q,r\},b) = \{q,r\}, \quad \Delta(\{p,r\},b) = \emptyset, \quad \Delta(\{p,q,r\},b) = \{q,r\}.$$

The converted DFA is shown below. The accessible states are shaded.



2. (a) Write a context-free grammar for the language $L_2 = \{\alpha \in \{a,b,c\}^* \mid \#a(\alpha) + \#b(\alpha) = \#c(\alpha)\}$. Here, $\#d(\alpha)$ means the number of occurrences of the symbol d in the string α, where $d \in \{a,b,c\}$. Write only the productions in your grammar, and mention which is the start symbol. **(5)**

Solution The following grammar with start symbol S generates L_2.

$$\begin{array}{rcl} S & \rightarrow & \epsilon \mid cST \mid TSc \mid SS\,, \\ T & \rightarrow & a \mid b\,. \end{array}$$

(b) Convert the grammar of Part (a) to Chomsky normal form. Show all the relevant steps briefly. **(5)**

Solution **Step 1:** Remove unit and ϵ productions.

There are no unit productions. The ϵ production $S \rightarrow \epsilon$ can be removed after adding the new rules $S \rightarrow cT$ and $S \rightarrow Tc$.

Step 2: Add non-terminal symbols for elements of Σ.

We add three non-terminal symbols A, B, C and the three rules $A \rightarrow a$, $B \rightarrow b$ and $C \rightarrow c$. We then rewrite the two rules $S \rightarrow cT$ and $S \rightarrow Tc$ respectively as $S \rightarrow CT$ and $S \rightarrow TC$. Moreover, $S \rightarrow cST$ is rewritten as $S \rightarrow CST$, and $S \rightarrow TSc$ as $S \rightarrow TSC$.

Step 3: Handle productions with more than two non-terminal symbols on the right sides.

We replace the rule $S \rightarrow CST$ by the two rules $S \rightarrow CU$ and $U \rightarrow ST$. Similarly, we replace the rule $S \rightarrow TSC$ by the two rules $S \rightarrow TV$ and $V \rightarrow SC$.

The grammar in Chomsky normal form, therefore, consists of the following rules. The non-terminal symbols A, B are redundant and not shown here.

$$\begin{array}{rcl} S & \rightarrow & CT \mid TC \mid CU \mid TV \mid SS\,, \\ U & \rightarrow & ST\,, \\ V & \rightarrow & SC\,, \\ T & \rightarrow & a \mid b\,, \\ C & \rightarrow & c\,. \end{array}$$

3. Let n be a fixed positive integer, and L_3 the set of all strings over $\{a, b\}$ in which the n-th last symbol is a. Prove that any NFA accepting L_3 must have at least $n+1$ states. (**Hint:** First, show that for any NFA $(Q, \{a, b\}, \Delta, S, F)$ accepting L_3, and for any string $\alpha \in \{a, b\}^*$, we have $\hat{\Delta}(S, \alpha) \neq \emptyset$.) **(5)**

Solution Let the NFA $N = (Q, \{a.b\}, \Delta, S, F)$ accept L_3. If $\hat{\Delta}(S, \alpha) = \emptyset$ for some α, we have $\hat{\Delta}(S, \alpha\beta) = \emptyset$ for any string β. For any choice of β of length n and starting with a, the string $\alpha\beta$ is in L_3 and must be accepted, so $\hat{\Delta}(S, \alpha\beta)$ must contain at least one final state and cannot be empty Therefore, $\hat{\Delta}(S, \alpha) \neq \emptyset$ for all strings α.

Now, suppose that N has $k \leqslant n$ states. There are 2^n strings of length n over $\{a, b\}$. Since $\hat{\Delta}(S, \alpha)$ can take at most $2^k - 1 \leqslant 2^n - 1$ possible values for any such string α, we must have two distinct strings α, β of length n such that $\hat{\Delta}(S, \alpha) = \hat{\Delta}(S, \beta)$. Let $i \in \{1, 2, \ldots, n\}$ be a position where α and β differ. By symmetry, we can assume that the i-th symbol in α is a, and the i-th symbol in β is b. Let γ be any string of length $i - 1$. Since $\hat{\Delta}(S, \alpha) = \hat{\Delta}(S, \beta)$, we must also have $\hat{\Delta}(S, \alpha\gamma) = \hat{\Delta}(S, \beta\gamma) = P$ (say). Since $\alpha\gamma \in L_3$, P must contain at least one final state. At the same time, $\beta\gamma \notin L_3$, so P must not contain any final state. This is a contradiction.

4. Using the pumping lemma, prove that $L_4 = \{a^m b^n \mid m, n \geqslant 1 \text{ and } \gcd(m, n) = 1\}$ is not regular. **(5)**

Solution Suppose that L_4 is regular, and let $k \geqslant 1$ be a pumping-lemma constant for L_4. Choose any prime $p > k$, and feed $\alpha = \epsilon$, $\beta = a^p$ and $\gamma = b^{(p+1)(p+2)\cdots(p+k)}$ to the pumping lemma. Since p is a prime larger than k, neither of $p + i$ is divisible by p for $i = 1, 2, \ldots, k$, so that $\gcd(p, (p+1)(p+2)\cdots(p+k)) = 1$, that is, $\alpha\beta\gamma \in L_4$. The pumping lemma supplies a decomposition of the form $\beta = \beta_1\beta_2\beta_3$ with $\beta_2 = a^t$ for some t in the range $1 \leqslant t \leqslant k$. Pumping in one occurrences of β_2 gives the string $\alpha\beta_1\beta_2^2\beta_3\gamma = a^{p+t}b^{(p+1)(p+2)\cdots(p+k)} \in L_4$. But $\gcd(p+t, (p+1)(p+2)\cdots(p+k)) = p + t \geqslant 3$ (since $p \geqslant 2$ and $t \geqslant 1$), a contradiction.

Roll no: ____________________ **Name:** ____

5. Let L_5 denote the non-regular set $\{a^n b^n \mid n \geqslant 0\}$. Prove or disprove the following two assertions.

(a) Any *infinite* subset of L_5 must be non-regular. **(5)**

Solution *True.* Let L' be an infinite subset of L_5. Suppose that L' is regular, and let k be a pumping-lemma constant for L'. Being infinite, L' contains a string longer than any $l \geqslant 0$. In particular, L' contains a string $a^n b^n$ with $n \geqslant k$, which we feed to the pumping lemma as $\alpha = \epsilon$, $\beta = a^n$ and $\gamma = b^n$. The pumping lemma supplies a decomposition $\beta = \beta_1\beta_2\beta_3$ with $|\beta_2| = t$, where $1 \leqslant t \leqslant k$. The lemma also guarantees that $\alpha\beta_1\beta_3\gamma = a^{n-t}b^n$ must be in L'. But $n - t \neq n$, so $a^{n-t}b^n$ cannot be in L_5. Finally, since $L' \subseteq L_5$, the string $a^{n-t}b^n$ cannot be in L' too, a contradiction.

(b) Any *infinite* subset of $\sim L_5 = \{a, b\}^* \setminus L_5$ must be non-regular. **(5)**

Solution *False.* For example, take the infinite subset $L'' = \{ab^n a \mid n \geqslant 0\}$ of $\sim L_5$. Since L'' is the language of the regular expression ab^*a, it is regular.