CSE322
Regular Expressions and their Identities

Lecture #6

Regular Expressions

We give a formal recursive definition of regular expressions over ∑ as follows:

1. Any terminal symbol (i.e. an element of ∑), Λ and Ø are regular expressions. When we view *a* in ∑ as a regular expression, we denote it by a.
2. The union of two regular expressions R1and R2 , written as R 1 + R2, is also a regular expression.
3. The concatenation of two regular expressions R1and R2, written as R1 R2, is also a regular expression.
4. The iteration (or closure) of a regular expression R written as R*, is also a regular expression.
5. If **R** is a regular expression, then **(R)** is also a regular expression.
6. The regular expressions over∑ are precisely those obtained recursively by the application of the rules *1-5* once or several times

Definitions

Any set represented by a regular expression is called a regular set.

If for example, *a, b* ε ∑. Then

(i) a denotes the set *{a},*

(ii) a + b denotes *{a, b},*

(iii) ab denotes *{ab},*

(iv) a* denotes the set {lambda, *a, aa. aaa,* ... } and

(v) (a + b)* denotes *{a, b}*.* The set represented by *R* is denoted by L(R),

Questions

Describe the following sets by regular expressions:

(a) {101}

(b) {abba}

(c) {01,10}

(d) {Λ ,ab}

(e) {abb,a, b, bba}

(f) {Λ , 0, 00, 000.... }

(g) {1, 11, 111 ... }

Solution

(a) Now. {1}. {O} are represented by 1 and O. respectively. 101 is obtained by concatenating 1,0 and 1 So. {10 l} is represented by 101.

(b) abba represents *{abba}.*

(c) As {01, 10} is the union of {01} and {10}, we have {01, 10} represented by 01 + 10

(d) The set {Λ , *ab} is represented by* Λ *+ ab*

(e) The set *{abb, a, b, bba} is represented by abb + a + b + bba.*

f) As {Λ , 0, 00, 000, ... } is simply {O}*. it is represented by 0*

(g) Any element in {1, 11, 111, ... } can be obtained by concatenating 1 and any element of {1}*. Hence 1(1)* represents {1, 11, 111, ... }

Problem

Describe the following sets by regular expressions:

(a) L_1 = the set of all strings of 0's and 1's ending in 00.

(b) L_2 = the set of all strings of 0's and 1's beginning with 0 and ending with 1.

(c) $L_3 = \{\Lambda, 11, 1111, 111111, \ldots\}$.

Solution

(a) Any string in L_1 is obtained by concatenating any string over {0, 1} and the string 00. {0, 1} is represented by **0 + 1**. Hence L_1 is represented by **(0 + 1)* 00**.

(b) As any element of L_2 is obtained by concatenating 0, any string over {0, 1} and 1, L_2 can be represented by **0(0 + 1)* 1**.

(c) Any element of L_3 is either Λ or a string of even number of 1's, i.e. a string of the form $(11)^n$, $n \geq 0$. So L_3 can be represented by **(11)***.

Solution

(a) Now. {1}. {O} are represented by 1 and O. respectively. 101 is obtained by concatenating 1,0 and 1 So. {10 I} is represented by 101.

(b) abba represents *{abba}.*

(c) As {01, 10} is the union of {01} and {10}, we have {01, 10} represented by 01 + 10

(d) The set {Λ , *ab} is represented by* Λ *+ ab*

(e) The set *{abb, a, b, bba} is represented by abb + a + b + bba.*

f) As {Λ , 0, 00, 000, ... } is simply {O}*. it is represented by 0*

(g) Any element in {1, 11, 111, ... } can be obtained by concatenating 1 and any element of {1}*. Hence 1(1)* represents {1, 11, 111, ... }

IDENTITIES FOR REGULAR EXPRESSIONS

I_1 $\quad \emptyset + \mathbf{R} = \mathbf{R}$

I_2 $\quad \emptyset\mathbf{R} = \mathbf{R}\emptyset = \emptyset$

I_3 $\quad \Lambda\mathbf{R} = \mathbf{R}\Lambda = \mathbf{R}$

I_4 $\quad \Lambda^* = \Lambda$ and $\emptyset^* = \Lambda$

I_5 $\quad \mathbf{R} + \mathbf{R} = \mathbf{R}$

I_6 $\quad \mathbf{R}^*\mathbf{R}^* = \mathbf{R}^*$

I_7 $\quad \mathbf{R}\mathbf{R}^* = \mathbf{R}^*\mathbf{R}$

I_8 $\quad (\mathbf{R}^*)^* = \mathbf{R}^*$

I_9 $\quad \Lambda + \mathbf{R}\mathbf{R}^* = \mathbf{R}^* = \Lambda + \mathbf{R}^*\mathbf{R}$

I_{10} $\quad (\mathbf{PQ})^*\mathbf{P} = \mathbf{P}(\mathbf{QP})^*$

I_{11} $\quad (\mathbf{P} + \mathbf{Q})^* = (\mathbf{P}^*\mathbf{Q}^*)^* = (\mathbf{P}^* + \mathbf{Q}^*)^*$

I_{12} $\quad (\mathbf{P} + \mathbf{Q})\mathbf{R} = \mathbf{PR} + \mathbf{QR}$ and $\mathbf{R}(\mathbf{P} + \mathbf{Q}) = \mathbf{RP} + \mathbf{RQ}$

THEORM

Theorem :

(Arden' s theorem) Let **P** and **Q** be two regular expressions over ∑. If P does not contain Λ, then the following equation in **R**, namely

R = Q + RP

has a unique solution (i.e. one and only one solution) given by **R = QP*.**

THEORM

Proof

$Q + (QP^*)P = Q(\Lambda + P^*P) = QP^*$ by I9

$$
\begin{aligned}
Q + RP &= Q + (Q + RP)P \\
&= Q + QP + RPP \\
&= Q + QP + RP^2 \\
&= Q + QP + QP^2 + \cdots + QP^i + RP^{i+1} \\
&= Q(\Lambda + P + P^2 + \cdots + P^i) + RP^{i+1}
\end{aligned}
$$

Important justification

$$\mathbf{R} = \mathbf{Q}(\Lambda + \mathbf{P} + \mathbf{P}^2 + \cdots + \mathbf{P}^i) + \mathbf{RP}^{i+1} \qquad \text{for } i \geq 0 \qquad (5.2)$$

We now show that any solution of (5.1) is equivalent to $\mathbf{QP}^*$. Suppose $\mathbf{R}$ satisfies (5.1), then it satisfies (5.2). Let w be a string of length i in the set $\mathbf{R}$. Then w belongs to the set $Q(\Lambda + \mathbf{P} + \mathbf{P}^2 + \ldots + \mathbf{P}^i) + \mathbf{RP}^{i+1}$. As $\mathbf{P}$ does not contain Λ, $\mathbf{RP}^{i+1}$ has no string of length less than $i + 1$ and so w is not in the set $\mathbf{RP}^{i+1}$. This means that w belongs to the set $Q(\Lambda + \mathbf{P} + \mathbf{P}^2 + \ldots + \mathbf{P}^i)$, and hence to $\mathbf{QP}^*$.

Consider a string w in the set $\mathbf{QP}^*$. Then w is in the set $\mathbf{QP}^k$ for some $k \geq 0$, and hence in $Q(\Lambda + \mathbf{P} + \mathbf{P}^2 + \cdots + \mathbf{P}^k)$. So w is on the R.H.S. of (5.2). Therefore, w is in $\mathbf{R}$ (L.H.S. of (5.2)). Thus $\mathbf{R}$ and $\mathbf{QP}^*$ represent the same set. This proves the uniqueness of the solution of (5.1). ∎

Note: Henceforth in this text, the regular expressions will be abbreviated r.e.

PROBLEM

(a) Give an r.e. for representing the set L of strings in which every 0 is immediately followed by at least two 1's.

(b) Prove that the regular expression $\mathbf{R} = \Lambda + \mathbf{1^*(011)^*(1^*\ (011)^*)^*}$ also describes the same set of strings.

SOLUTION

Solution

(a) If w is in L, then either (a) w does not contain any 0, or (b) it contains a 0 preceded by 1 and followed by 11. So w can be written as $w_1 w_2 \ldots w_n$, where each w_i is either 1 or 011. So L is represented by the r.e. $(\mathbf{1} + \mathbf{011})^*$.

(b) $\mathbf{R} = \Lambda + \mathbf{P}_1\mathbf{P}_1^*$, where $\mathbf{P}_1 = \mathbf{1}^*(\mathbf{011})^*$

$= \mathbf{P}_1^*$ using I_9

$= (\mathbf{1}^*(\mathbf{011})^*)^*$

$= (\mathbf{P}_2^*\mathbf{P}_3^*)^*$ letting $\mathbf{P}_2 = \mathbf{1}$, $\mathbf{P}_3 = \mathbf{011}$

$= (\mathbf{P}_2 + \mathbf{P}_3)^*$ using I_{11}

$= (\mathbf{1} + \mathbf{011})^*$

PROBLEM & SOLUTION

Prove $(1 + 00^*1) + (1 + 00^*1)(0 + 10^*1)^* (0 + 10^*1) = 0^*1(0 + 10^*1)^*$.

Solution

$$\text{L.H.S.} = (1 + 00^*1)(\Lambda + (0 + 10^*1)^* (0 + 10^*1)\Lambda \qquad \text{using } I_{12}$$

$$= (1 + 00^*1)(0 + 10^*1)^* \qquad \text{using } I_9$$

$$= (\Lambda + 00^*)1(0 + 10^*1)^* \qquad \text{using } I_{12} \text{ for } 1 + 00^*1$$

$$= 0^*1(0 + 10^*1)^* \qquad \text{using } I_9$$

$$= \text{R.H.S.}$$

EXAMPLE 5.34

Prove that $P + PQ^*Q = a^*bQ^*$ where $P = b + aa^*b$ and Q is any regular expression.

Proof

$$
\begin{aligned}
\text{L.H.S.} &= \mathbf{P}\Lambda + \mathbf{PQ^*Q} && \text{by } I_3\\
&= \mathbf{P}(\Lambda + \mathbf{Q^*Q}) && \text{by } I_{12}\\
&= \mathbf{PQ^*} && \text{by } I_9\\
&= (\mathbf{b} + \mathbf{aa^*b})\mathbf{Q^*} && \text{by definition of } \mathbf{P}\\
&= (\Lambda\mathbf{b} + \mathbf{aa^*b})\mathbf{Q^*} && \text{by } I_3\\
&= (\Lambda + \mathbf{aa^*})\mathbf{bQ^*} && \text{by } I_{12}\\
&= \mathbf{a^*bQ^*} && \text{by } I_9\\
&= \text{R.H.S.}
\end{aligned}
$$

Difference between Null and Ø

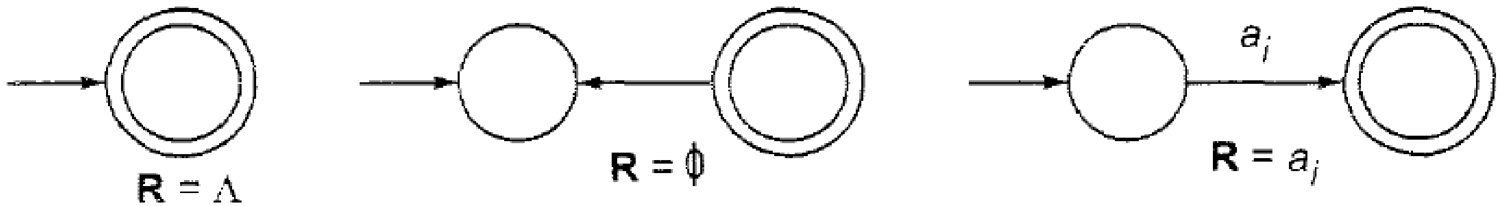


Fig. 5.5 Transition systems for recognizing elementary regular sets.

TRANSITION SYSTEM CONTAINING Λ -MOVES

Suppose we want to replace a Λ -move from vertex *V1 to vertex V2' Then* we proceed as follows:

Step 1- Find all the edges starting from *V2.*

Step 2- Duplicate all these edges starting from *V1' without changing the edge* labels.

Step 3- If V1 is an initial state, make V2 also as initial state.

Step 4- If V2 is a final state. make V1 also as the final state.

EXAMPLE

Consider a finite automaton, with Λ-moves, given in Fig. 5.1. Obtain an equivalent automaton without Λ-moves.

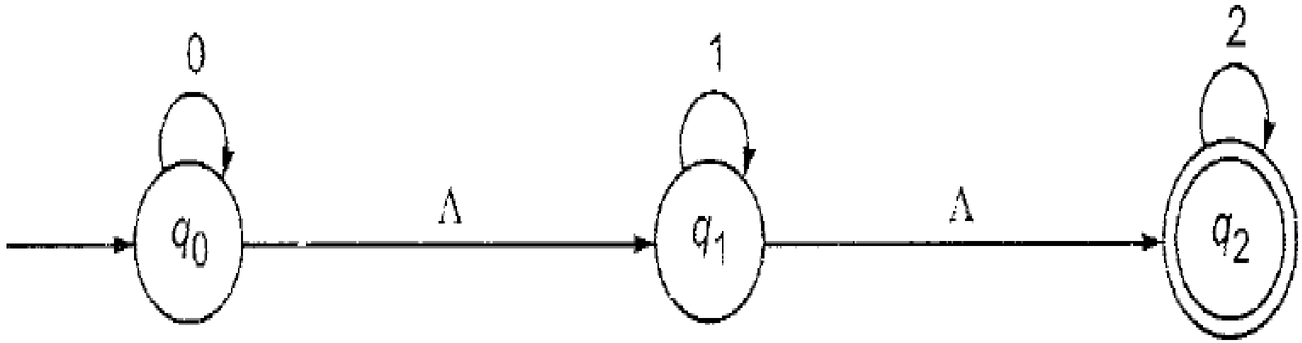
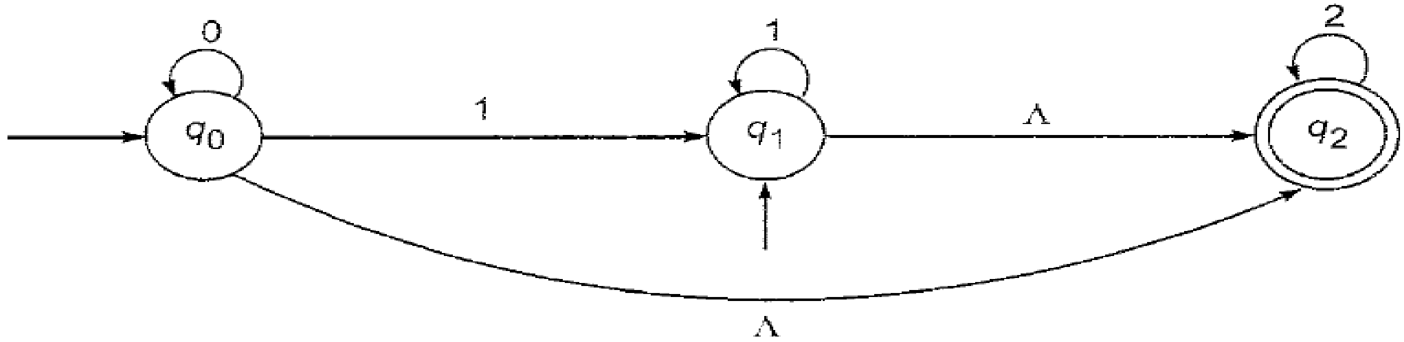
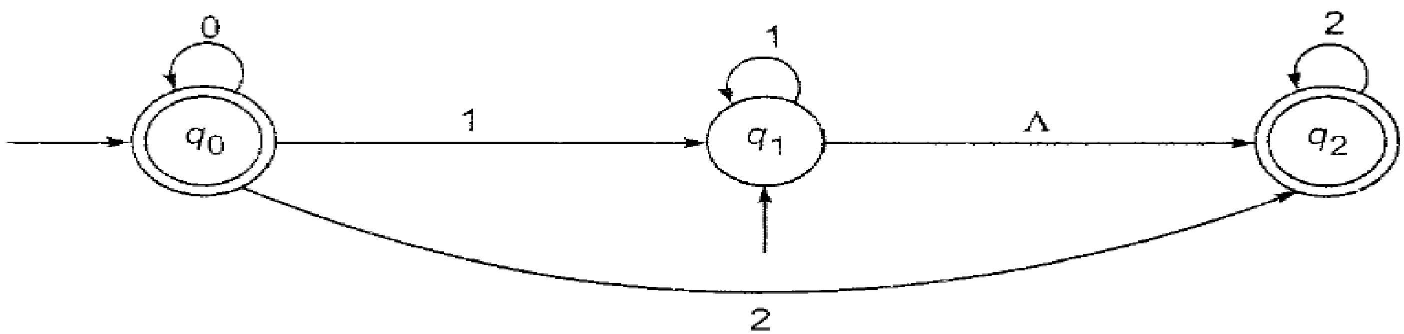


Fig. 5.1 Finite automaton of Example 5.5.

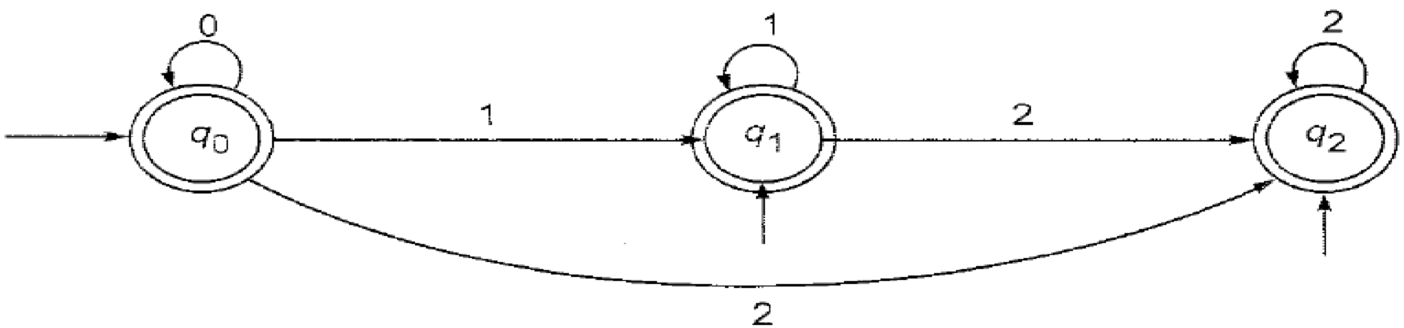
EXAMPLE



(a)



(b)



(c)

Construct RE from given FA

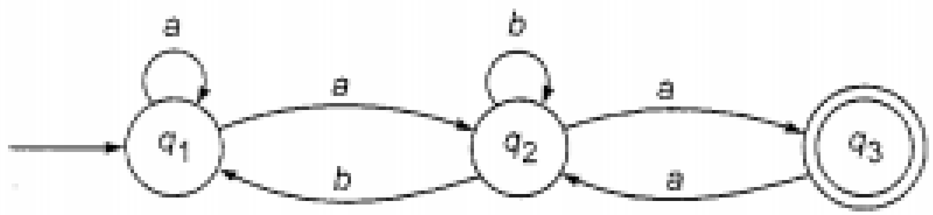
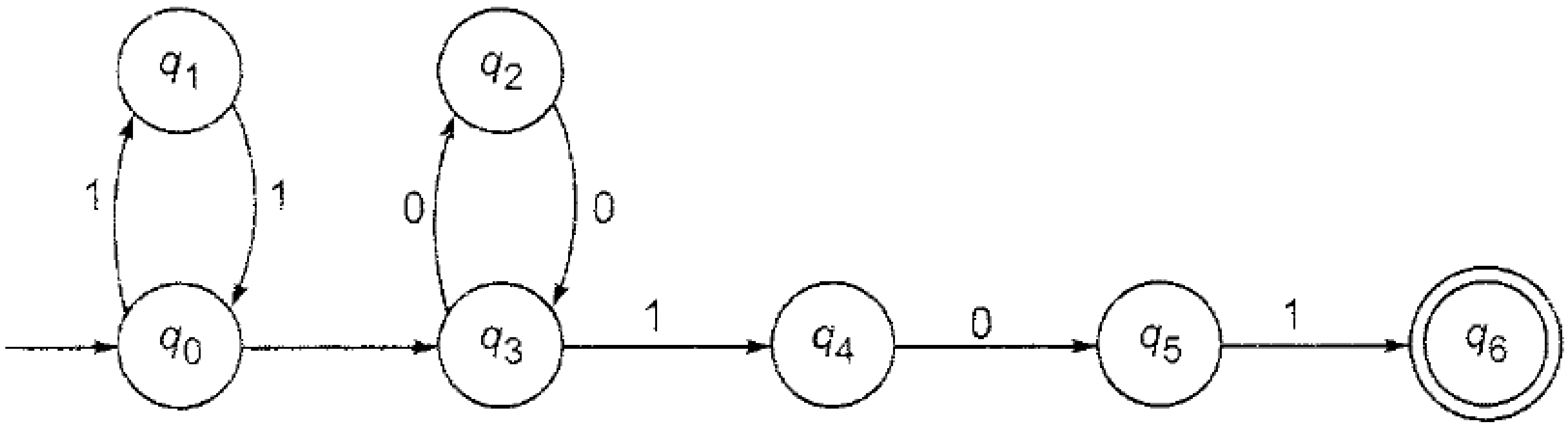


Fig. 5.13 Transition system of Example 5.8.

Convert this automata with null moves to automata without it



Obtain the deterministic graph (system) equivalent to the transition system given in Fig. 5.11.

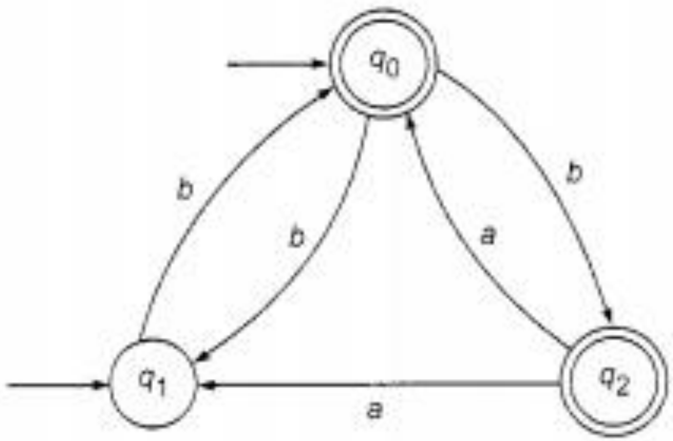


Fig. 5.11 Nondeterministic transition system of Example 5.7.

5.2.4 ALGEBRAIC METHOD USING ARDEN'S THEOREM

The following method is an extension of the Arden's theorem (Theorem 5.1). This is used to find the r.e. recognized by a transition system.

The following assumptions are made regarding the transition system:

(i) The transition graph does not have Λ-moves.
(ii) It has only one initial state, say v_1.
(iii) Its vertices are $v_1 \ldots v_n$.
(iv) $\mathbf{V}_i$ the r.e. represents the set of strings accepted by the system even though v_i is a final state.
(v) α_{ij} denotes the r.e. representing the set of labels of edges from v_i to v_j. When there is no such edge, $\alpha_{ij} = \emptyset$. Consequently, we can get the following set of equations in $\mathbf{V}_1 \ldots \mathbf{V}_n$:

$$\mathbf{V}_1 = \mathbf{V}_1\alpha_{11} + \mathbf{V}_2\alpha_{21} + \cdots + \mathbf{V}_n\alpha_{n1} + \Lambda$$

$$\mathbf{V}_2 = \mathbf{V}_1\alpha_{12} + \mathbf{V}_2\alpha_{22} + \cdots + \mathbf{V}_n\alpha_{n2}$$

$$\vdots$$

$$\mathbf{V}_n = \mathbf{V}_1\alpha_{1n} + \mathbf{V}_2\alpha_{2n} + \cdots + \mathbf{V}_n\alpha_{nn}$$

By repeatedly applying substitutions and Theorem 5.1 (Arden's theorem), we can express $\mathbf{V}_i$ in terms of α_{ij}'s.

For getting the set of strings recognized by the transition system, we have to take the 'union' of all $\mathbf{V}_i$'s corresponding to final states.

Construct RE from Finite Automata

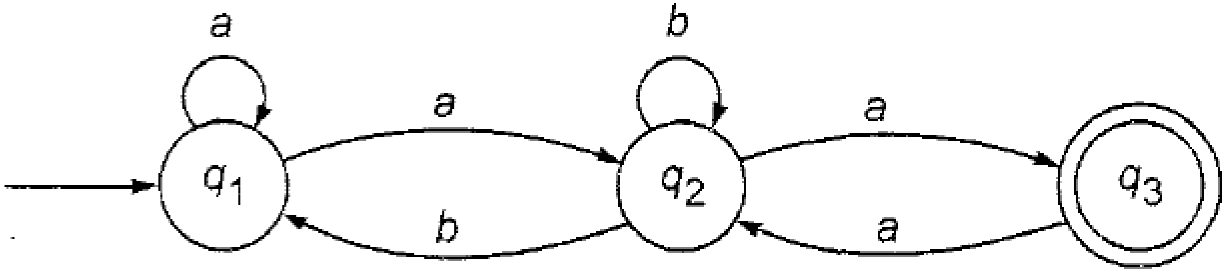


Fig. 5.13 Transition system of Example 5.8.

answer

- $(a + a(b + aa)^*b)^*a(b + aa)^*a$

Construct RE from given DFA

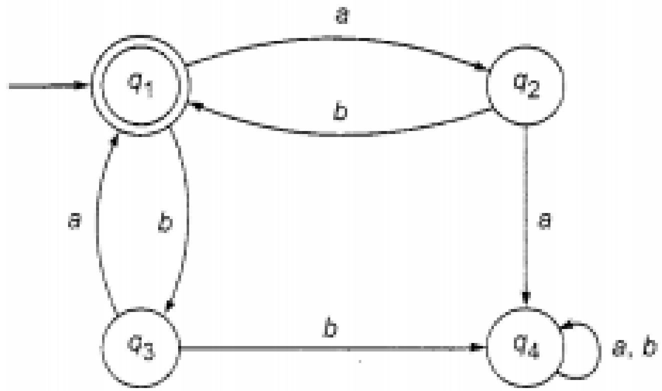


Fig. 5.14 Finite automaton of Example 5.9.

Answer

- (ab + ba)*

Construct RE from given DFA, take q1 as initial state

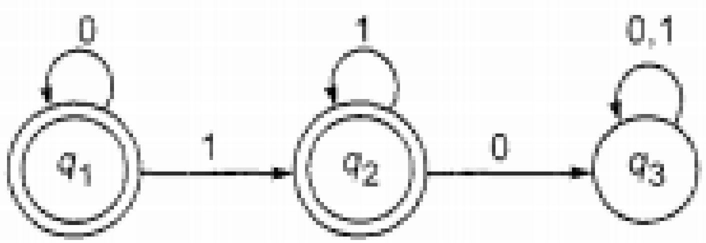


Fig. 5.15 Finite automaton of Example 5.10.

Answer

- 0*(1*)

Construct RE from given DFA

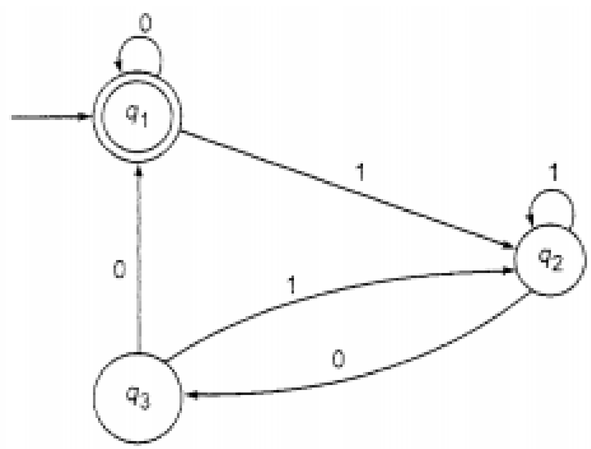


Fig. 5.16 Finite automaton of Example 5.11.

Answer

- $(0 + 1(1 + 01)^* 00)^*$

Construct RE from given DFA

- Q1 as initial state

EXAMPLE 5.12

Find the regular expression corresponding to Fig. 5.17.

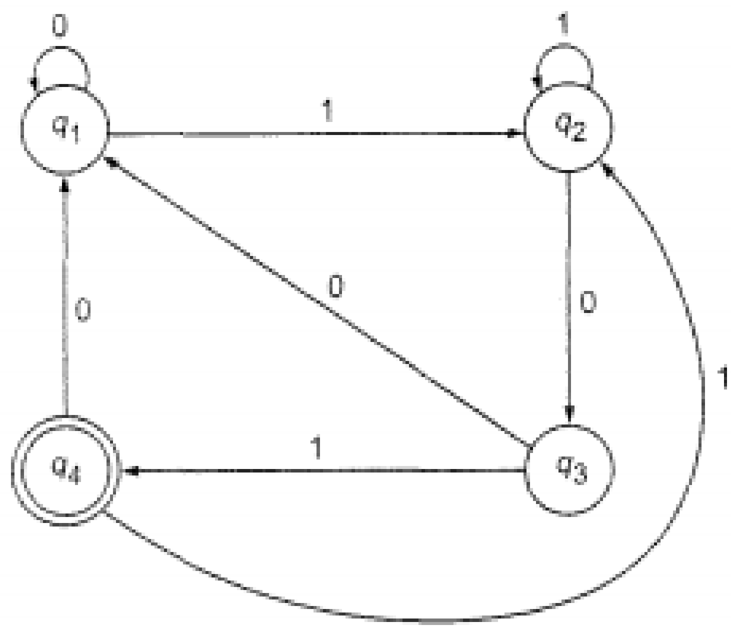


Fig. 5.17 Finite automaton of Example 5.12.

- $(0 + 1(1 + 011)^*(00 + 010)^*(1(1 + 011)^* 01)$

5.2.6 EQUIVALENCE OF TWO FINITE AUTOMATA

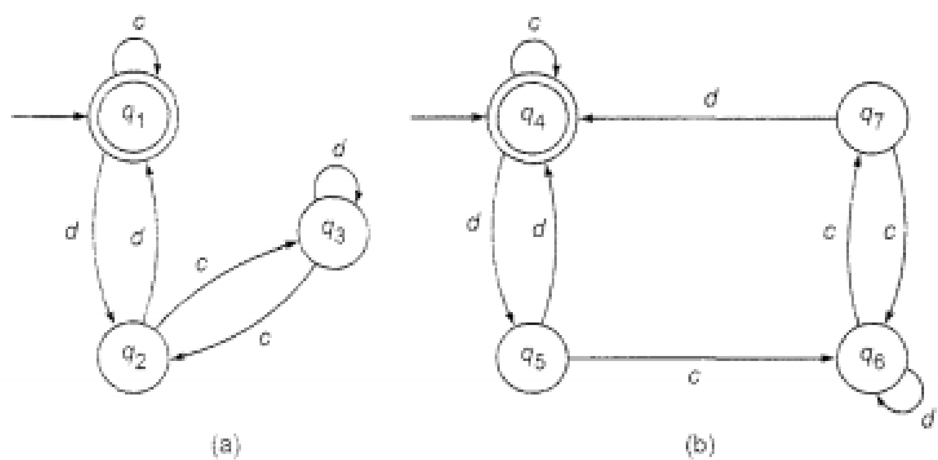


Fig. 5.23 (a) Automaton M and (b) automaton M'.

EXAMPLE 5.16

Show that the automata M_1 and M_2 defined by Fig. 5.24 are not equivalent.

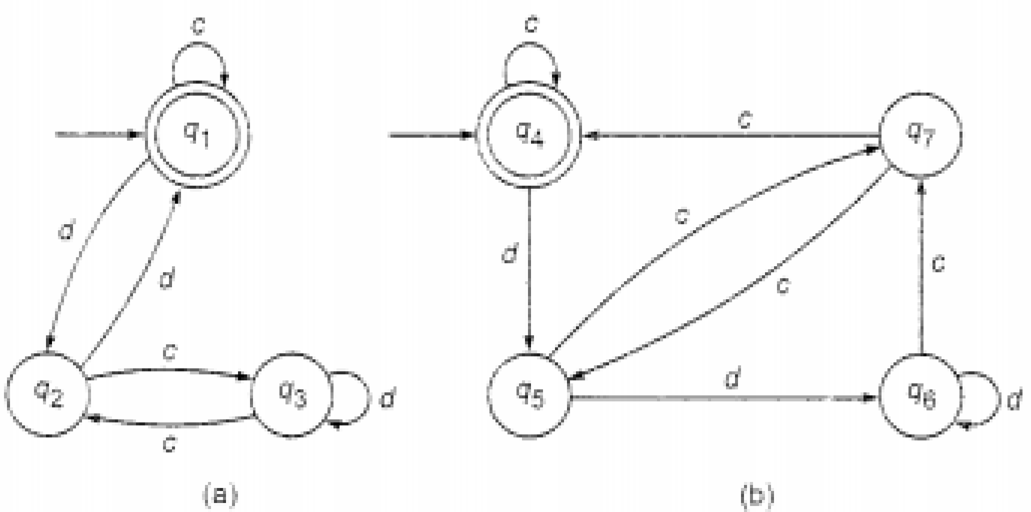


Fig. 5.24 (a) Automaton M_1 and (b) automaton M_2.

Closure properties of regular sets

- **Property 1**. *The union of two regular set is regular.*
- **Proof** −
- Let us take two regular expressions
- $RE_1 = a(aa)^*$ and $RE_2 = (aa)^*$
- So, $L_1 = \{a, aaa, aaaaa, ...\}$ (Strings of odd length excluding Null)
- and $L_2 = \{ \varepsilon, aa, aaaa, aaaaaa, ...\}$ (Strings of even length including Null)
- $L_1 \cup L_2 = \{ \varepsilon, a, aa, aaa, aaaa, aaaaa, aaaaaa, ...\}$
- (Strings of all possible lengths including Null)
- $RE(L_1 \cup L_2) = a^*$ (which is a regular expression itself)
- **Hence, proved.**

Closure properties of regular sets

- **Property 2.** *The intersection of two regular set is regular.*
- **Proof** −
- Let us take two regular expressions
- $RE_1 = a(a^*)$ and $RE_2 = (aa)^*$
- So, L_1 = { a,aa, aaa, aaaa, ...} (Strings of all possible lengths excluding Null)
- L_2 = { ε, aa, aaaa, aaaaaa,.......} (Strings of even length including Null)
- $L_1 \cap L_2$ = { aa, aaaa, aaaaaa,.......} (Strings of even length excluding Null)
- RE $(L_1 \cap L_2) = aa(aa)^*$ which is a regular expression itself.
- **Hence, proved.**

Closure properties of regular sets

- **Property 3.** *The complement of a regular set is regular.*
- **Proof** −
- Let us take a regular expression −
- RE = (aa)*
- So, L = {ε, aa, aaaa, aaaaaa, ...} (Strings of even length including Null)
- Complement of **L** is all the strings that is not in **L**.
- So, L' = {a, aaa, aaaaa, ...} (Strings of odd length excluding Null)
- RE (L') = a(aa)* which is a regular expression itself.
- **Hence, proved.**

Closure properties of regular sets

- **Property 4.** *The difference of two regular set is regular.*
- **Proof** −
- Let us take two regular expressions −
- RE_1 = a (a*) and RE_2 = (aa)*
- So, L_1 = {a, aa, aaa, aaaa, ...} (Strings of all possible lengths excluding Null)
- L_2 = { ε, aa, aaaa, aaaaaa,.......} (Strings of even length including Null)
- $L_1 – L_2$ = {a, aaa, aaaaa, aaaaaaa, ...}
- (Strings of all odd lengths excluding Null)
- RE ($L_1 – L_2$) = a (aa)* which is a regular expression.

Closure properties of regular sets

- **Property 5.** *The reversal of a regular set is regular.*
- **Proof** −
- We have to prove $\mathbf{L^R}$ is also regular if **L** is a regular set.
- Let, L = {01, 10, 11, 10}
- RE (L) = 01 + 10 + 11 + 10
- L^R = {10, 01, 11, 01}
- RE (L^R) = 01 + 10 + 11 + 10 which is regular
- **Hence, proved.**
- **Property 6.** *The closure of a regular set is regular.*
- **Proof** −
- If L = {a, aaa, aaaaa, ...} (Strings of odd length excluding Null)
- i.e., RE (L) = a (aa)*
- L* = {a, aa, aaa, aaaa , aaaaa,……………} (Strings of all lengths excluding Null)
- RE (L*) = a (a)*
- **Hence, proved.**

Closure properties of regular sets

- **Property 7.** *The concatenation of two regular sets is regular.*
- **Proof –**
- Let $RE_1 = (0+1)^*0$ and $RE_2 = 01(0+1)^*$
- Here, L_1 = {0, 00, 10, 000, 010, ...} (Set of strings ending in 0)
- and L_2 = {01, 010,011,.....} (Set of strings beginning with 01)
- Then, $L_1 L_2$ = {001,0010,0011,0001,00010,00011,1001,10010,.............}
- Set of strings containing 001 as a substring which can be represented by an RE − $(0 + 1)^*001(0 + 1)^*$

Construct DFA Equivalent to regular expression

- (0 +1)*(00 + 11)(0 + 1)*
- 10 + (0 + 11)0*1

Equivalence of 2 regular expression

- Prove $(a + b)^* = a^*(ba^*)^*$ by constructing regular expressions

CONSTRUCTION OF A REGULAR GRAMMAR GENERATING *T(M) FOR A GIVEN DFA M*

- Construct a regular grammar G generating the regular set represented by
- $P = a^*b(a + b)^*$.

CONSTRUCTION OF A TRANSITION SYSTEM *M* ACCEPTING *L(G) FOR A GIVEN REGULAR* GRAMMAR G

- Let G =*({Ao, A1,, {a. b}. P, Ao).* where *P* consists of Ao->aA1,A1->bA1,A1->a,A1->bA0
- Construct a transition system *M accepting L(G).*

5.4 APPLICATION OF PUMPING LEMMA

This theorem can be used to prove that certain sets are not regular. We now give the steps needed for proving that a given set is not regular.

Step 1 Assume that L is regular. Let n be the number of states in the corresponding finite automaton.

Step 2 Choose a string w such that $|w| \geq n$. Use pumping lemma to write $w = xyz$, with $|xy| \leq n$ and $|y| > 0$.

Step 3 Find a suitable integer i such that $xy^iz \notin L$. This contradicts our assumption. Hence L is not regular.

Note: The crucial part of the procedure is to find i such that $xy^iz \notin L$. In some cases we prove $xy^iz \notin L$ by considering $|xy^iz|$. In some cases we may have to use the 'structure' of strings in L.