UNIT-II

Subject-Theory of Computation

Prof. Shweta Tiwaskar

Shweta.tiwaskar@viit.ac.in

Department of Computer Engineering

VISHWAKARMA INSTITUTES VI

BRACT'S, Vishwakarma Institute of Information Technology, Pune-48

(An Autonomous Institute affiliated to Savitribai Phule Pune University)

(NBA and NAAC accredited, ISO 9001:2015 certified)

Regular Expression

1. Any terminal symbol (i.e. an element of Σ), Λ and $\emptyset$ are regular expressions. When we view a in Σ as a regular expression, we denote it by $\mathbf{a}$.
2. The union of two regular expressions $\mathbf{R}_1$ and $\mathbf{R}_2$, written as $\mathbf{R}_1 + \mathbf{R}_2$, is also a regular expression.
3. The concatenation of two regular expressions $\mathbf{R}_1$ and $\mathbf{R}_2$, written as $\mathbf{R}_1\mathbf{R}_2$, is also a regular expression.
4. The iteration (or closure) of a regular expression $\mathbf{R}$, written as $\mathbf{R}^*$, is also a regular expression.
5. If $\mathbf{R}$ is a regular expression, then $(\mathbf{R})$ is also a regular expression.
6. The regular expressions over Σ are precisely those obtained recursively by the application of the rules 1–5 once or several times.

- Regular Set: Any set represented by a regular expression is called regular set

Application of Regular Expression(RE)

- RE are useful for numerous practical day to day tasks that a data scientist encounters. They are used everywhere
- wide variety of text /data processing tasks
- natural language processing
- pattern matching
- Data validation
- Data scraping (especially web scraping)
- Data extraction
- Simple parsing, the production of syntax highlighting systems, and many other tasks.

Regular Expression	Regular Set
a	{a}
(a+b)	{a,b}
ab	{ab}
a*	{ λ , a, aa, aaa, aaaa,………}
a+ or a(a)*	{a, aa, aaa, aaaa,………}
(0+1)*	{ null, 0,1,01,10,11,00,……}
(a+b)*	{null,a,b,ab,aa,bb,ba,….}
(a+b)* ab	{ab,aab,bab,aab,…..}
(aa)*	{null,aa,aaaa,aaaaaa……}
(0*10*)	{ 1,01,10, 010,00100 ,….. }
(aa)*(bb)*b	{ b, aab, bbb, aabbb…… }
(aa + ab + ba + bb)*	{ null, aa,ab,ba,bb,aaab,……. }
(0+€) (1+€)	{ null,0,1,01 }
(0+10*)	{0, 1,10,100,1000…….}

Regular SET	Regular Expression
{101}	101
{abba}	abba
{01,10}	01+10
{λ,ab}	$(\lambda + ab)$
{1,11,111,1111,........}	$1^+ = 1(1)^*$

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

{00,000,100,0000,1100.....}

a) Answer: $(0+1)^*00$

{01,001,010,........}

B) Answer:0 $(0+1)^*1$

c) $(11)^*$

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

Answer=(1+011)*

The set of all strings over $\{a, b\}$ with three consecutive b's.
The set of all strings over $\{0, 1\}$ beginning with 00.
The set of all strings over $\{0, 1\}$ ending with 00 and beginning with 1.

Answers:

1: (a+b)* bbb (a+b)*

2. 00 (0+1)*

3. 1(0+1)*00

Identities for Regular Expression

Two regular expressions **P** and **Q** are equivalent (we write **P** = **Q**) if **P** and **Q** represent the same set of strings.

We now give the identities for regular expressions; these are useful for simplifying 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 $\quad \Lambda + \mathbf{R}\mathbf{R}^* = \mathbf{R}^* = \Lambda + \mathbf{R}^*\mathbf{R}$

Identities for Regular Expression

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

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

(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.

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

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

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

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

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

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

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

HW

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

Prove $(\mathbf{1} + \mathbf{00^*1}) + (\mathbf{1} + \mathbf{00^*1})(\mathbf{0} + \mathbf{10^*1})^* (\mathbf{0} + \mathbf{10^*1}) = \mathbf{0^*1}(\mathbf{0} + \mathbf{10^*1})^*$.

Solution

$$
\begin{aligned}
\text{L.H.S.} &= (\mathbf{1} + \mathbf{00^*1})(\Lambda + (\mathbf{0} + \mathbf{10^*1})^* (\mathbf{0} + \mathbf{10^*1})\Lambda && \text{using } I_{12} \\
&= (\mathbf{1} + \mathbf{00^*1})(\mathbf{0} + \mathbf{10^*1})^* && \text{using } I_9 \\
&= (\Lambda + \mathbf{00^*})\mathbf{1}(\mathbf{0} + \mathbf{10^*1})^* && \text{using } I_{12} \text{ for } \mathbf{1} + \mathbf{00^*1} \\
&= \mathbf{0^*1}(\mathbf{0} + \mathbf{10^*1})^* && \text{using } I_9 \\
&= \text{R.H.S.}
\end{aligned}
$$

Prove $(\mathbf{1} + \mathbf{00^*1}) + (\mathbf{1} + \mathbf{00^*1})(\mathbf{0} + \mathbf{10^*1})^* (\mathbf{0} + \mathbf{10^*1}) = \mathbf{0^*1}(\mathbf{0} + \mathbf{10^*1})^*$.

Solution

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

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

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

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

= R.H.S.

Arden's Theorem

- Let **P** and **Q** be two regular expressions.
- If **P** does not contain null string, then **R = Q + RP** has a unique solution that is **R = QP***
- **Proof** −
- **R = Q + RP**

R = Q + (Q + RP)P [After putting the value R = Q + RP]

R= Q + QP + RPP

When we put the value of **R** recursively again and again, we get the following equation −

$R = Q + QP + QP^2 + QP^3$.....

$R = Q (\varepsilon + P + P^2 + P^3 + ...)$

R = QP* [As P* represents (ε + P + P2 + P3 + ….)]

Hence, proved.

Algebric Method Using Arden's Theorem

- The transition graph does not have NULL moves.
- It has only one initial state ie. V1
- Vertices are V1….Vn.
- $\mathbf{R_{ij}}$ represents the set of labels of edges from $\mathbf{V_i}$ to $\mathbf{V_j}$, if no such edge exists, then $\mathbf{R_{ij} = \emptyset}$
- **Step 1** − Create equations as the following form for all the states of the DFA having n states with initial state V_1.

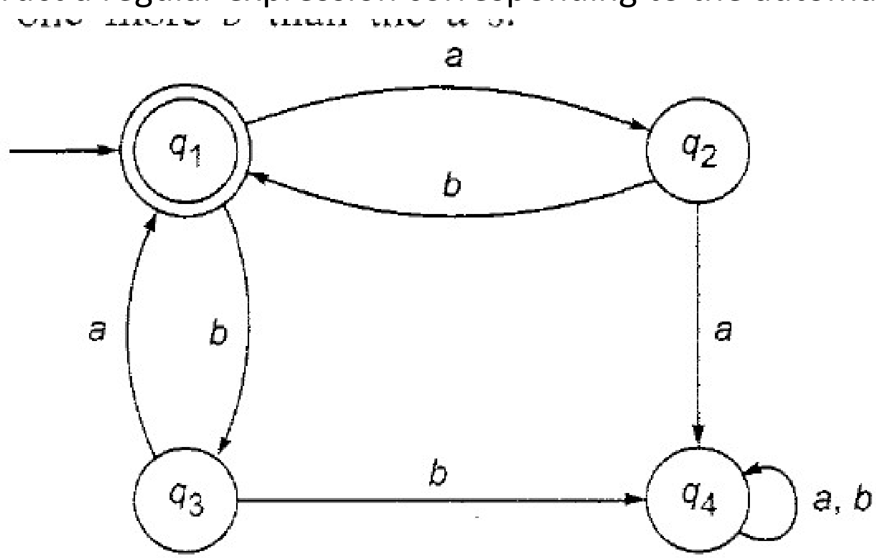
  $V_1 = V_1R_{11} + V_2R_{21} + \ldots + V_nR_{n1} + \varepsilon$

  $V_2 = V_1R_{12} + V_2R_{22} + \ldots + V_nR_{n2}$

  …………………………

  $V_n = V_1R_{1n} + V_2R_{2n} + \ldots + V_nR_{nn}$
- **Step 2** − Solve these equations to get the equation for the final state in terms of $\mathbf{R_{ij}}$

Construct a regular expression corresponding to the automata given below using Arden’s Theorem.

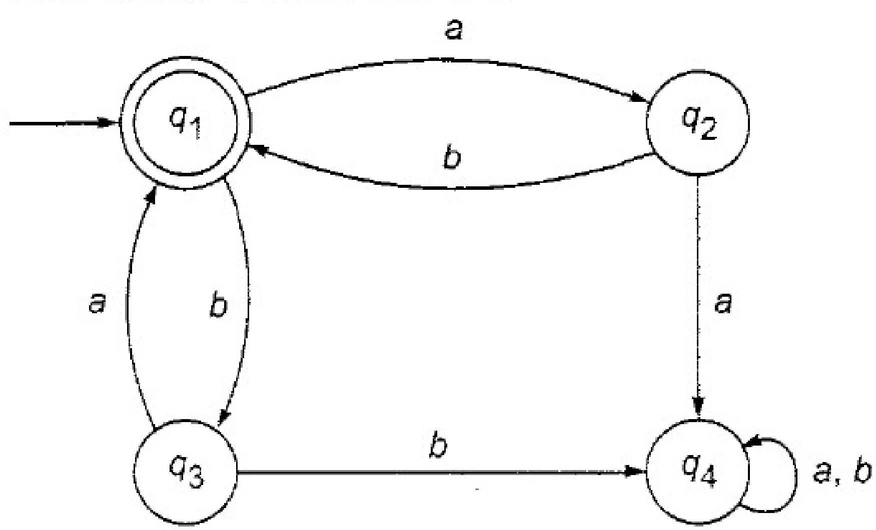


q1=q2b+q3a+null
q2=q1a
q3=q1b
q4=q2a+q3b+q4a+q4b

Put eq. q2 7 q3 in eq q1
q1= q1ab+q1ba+null
q1=q1(ab+ba)+null
R=Q=RP R=QP* R=q1, Q=null, P= (ab+ba)
q1=null (ab+ba)*=

(ab+ba)*---Answer

Construct a regular expression corresponding to the automata given below using Arden’s Theorem.



q1=q2b+q3 a+λ

q2=q1a
q3=q1b
q4=q2a+q3b+q4a+q4b
Put eq. of q2 and q3 in q1
q1=q1ab+q1ba+ λ

q1=q1(ab+ba)+^

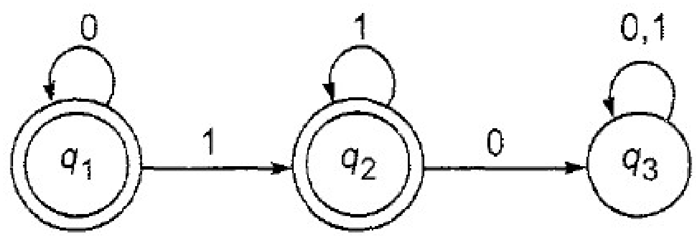
R=Q+RP= QP*

R=q1, Q=null, P=(ab+ba)
q1=null(ab+ba)*

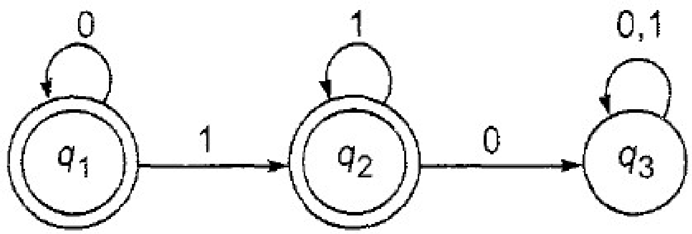
q1=(ab+ba)*

HW

Describe in English the set accepted by the finite automaton whose transition diagram is as shown in Fig. 5.15.



Describe in English the set accepted by the finite automaton whose transition diagram is as shown in Fig. 5.15.



Solution

We can apply the above method directly as the transition diagram does not contain more than one initial state and there are no Λ-moves. We get the following equations for $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$.

$$\mathbf{q}_1 = \mathbf{q}_1\mathbf{0} + \Lambda$$

$$\mathbf{q}_2 = \mathbf{q}_1\mathbf{1} + \mathbf{q}_2\mathbf{1}$$

$$\mathbf{q}_3 = \mathbf{q}_2\mathbf{0} + \mathbf{q}_3(\mathbf{0} + \mathbf{1})$$

By applying Theorem 5.1 to the $\mathbf{q}_1$-equation, we get

$$\mathbf{q}_1 = \Lambda\mathbf{0}^* = \mathbf{0}^*$$

So,

$$\mathbf{q}_2 = \mathbf{q}_1\mathbf{1} + \mathbf{q}_2\mathbf{1} = \mathbf{0}^*\mathbf{1} + \mathbf{q}_2\mathbf{1}$$

Therefore,

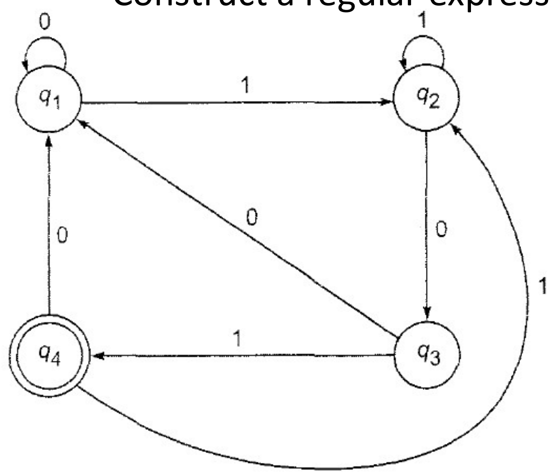
$$\mathbf{q}_2 = (\mathbf{0}^*\mathbf{1})\mathbf{1}^*$$

As the final states are $\mathbf{q}_1$ and $\mathbf{q}_2$, we need not solve for $\mathbf{q}_3$:

$$\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{0}^* + \mathbf{0}^*(\mathbf{1}\mathbf{1}^*) = \mathbf{0}^*(\Lambda + \mathbf{1}\mathbf{1}^*) = \mathbf{0}^*(\mathbf{1}^*) \qquad \text{by } I_9$$

The strings represented by the transition graph are **0*1***. We can interpret the strings in the English language in the following way: The strings accepted by the finite automaton are precisely the strings of any number of 0's (possibly Λ) followed by a string of any number of 1's (possibly Λ).

Construct a regular expression for the following diagram



Solution

There is only one initial state, and there are no Λ-moves. So, we form the equations corresponding to $\mathbf{q_1}$, $\mathbf{q_2}$, $\mathbf{q_3}$, $\mathbf{q_4}$:

$$\mathbf{q_1 = q_10 + q_30 + q_40 + \Lambda}$$

$$\mathbf{q_2 = q_11 + q_21 + q_41}$$

$$\mathbf{q_3 = q_20}$$

$$\mathbf{q_4 = q_31}$$

Now,

$$\mathbf{q_4 = q_31 = (q_20)1 = q_201}$$

Thus, we are able to write $\mathbf{q_3}$, $\mathbf{q_4}$ in terms of $\mathbf{q_2}$. Using the $\mathbf{q_2}$-equation, we get

$$\mathbf{q_2 = q_11 + q_21 + q_2011 = q_11 + q_2(1 + 011)}$$

By applying Theorem 5.1, we obtain

$$\mathbf{q}_2 = (\mathbf{q}_1\mathbf{1})(\mathbf{1} + \mathbf{011})^* = \mathbf{q}_1(\mathbf{1}(\mathbf{1} + \mathbf{011})^*)$$

From the $\mathbf{q}_1$-equation, we have

$$\begin{aligned}\mathbf{q}_1 &= \mathbf{q}_1\mathbf{0} + \mathbf{q}_2\mathbf{00} + \mathbf{q}_2\mathbf{010} + \Lambda \\ &= \mathbf{q}_1\mathbf{0} + \mathbf{q}_2(\mathbf{00} + \mathbf{010}) + \Lambda \\ &= \mathbf{q}_1\mathbf{0} + \mathbf{q}_1\mathbf{1}(\mathbf{1} + \mathbf{011})^* (\mathbf{00} + \mathbf{010}) + \Lambda\end{aligned}$$

Again, by applying Theorem 5.1, we obtain

$$\mathbf{q}_1 = \Lambda(\mathbf{0} + \mathbf{1}(\mathbf{1} + \mathbf{011})^* (\mathbf{00} + \mathbf{010}))^*$$

$$\begin{aligned}\mathbf{q}_4 &= \mathbf{q}_2\mathbf{01} = \mathbf{q}_1\mathbf{1}(\mathbf{1} + \mathbf{011})^* \mathbf{01} \\ &= (\mathbf{0} + \mathbf{1}(\mathbf{1} + \mathbf{011})^*(\mathbf{00} + \mathbf{010}))^*(\mathbf{1}(\mathbf{1} + \mathbf{011})^* \mathbf{01})\end{aligned}$$