UNIT-2

Regular expressions (RE):

regular language consist of a generator : Regular grammar
acceptor : FSM
reprasentor : RE

Def: RE are useful for reprasenting certain set of strings in an algebric fashion actualy these describes the languages accepted by finite state automata.

If R_1 and R_2 are 2 regular expressions then

$R_1 + R_2$ - union is possible

R_1^* - possible

R_1^+ - possible

$R_1 \cdot R_2$ - concatination is possible

× Any set represented by RE is called regular set

Ex:

Q. $\Sigma = \{a,b\}$ then (a) a denotes set $\{a\}$

b) $a+b$ denotes set $\{a,b\}$

c) ab denotes set $\{ab\}$

d) a^* denotes set $\{\epsilon, a, aa, aaa \ldots\}$

e) $(a+b)^*$ denotes set $\{a,b\}^*$ (or)

$\{\epsilon, a, b, ab, ba, aba \ldots\}$

Q. Describe following sets by RE

a) $\{101\}$ c) $\{01, 10\}$ e) $\{abb, a, b, bba\}$

b) $\{abba\}$ d) $\{\epsilon, ab\}$ f) $\{\epsilon, 0, 00, 000\}$ g) $\{1, 11, 111, \ldots\}$

a) 101

b) ~~xxx~~ $abba$

c) $01+10$

d) $\epsilon + ab$

e) $abb + a + b + bba$

f) $\epsilon + 0 + 00 + 000$

$= 0^*$

g) 1^+ (or)

$1.\{\epsilon, 1, 11, 111 \ldots\}$

Q) Describe following sets by RE

a) L1 = Set of all strings of 0's and 1's ending with '00'

b) L2 = Set of all strings of 0's and 1's beginning with 0 and ending with 1.

c) $L3 = \{\epsilon, 11, 1111, 111111, \ldots\}$

a) _____ 00

$= (0+1)^* 00$

Any string in L1 is obtained by concatinating any string over $\{0,1\}$ and string ending with 00

Hence L1 is represented as

b) 0 ____ 1

$= 0(0+1)^* 1$

c) $(11)^*$

Identities for RE / Algebraic expressions

$I_1 : \phi + R = R$

$I_2 : \phi R + R\phi = \phi$

$I_3 : \epsilon R + R\epsilon = R$

$I_4 : \epsilon^* = \epsilon$ & $\phi^* = \epsilon$

$I_5 : R + R = R$

$I_6 : R^* R^* = R^*$

$I_7 : RR^* = R^* R$

$I_8 : (R^*)^* = R^*$

* $I_9 : \epsilon + RR^* = R^* = \epsilon + R^* R$

* $I_{10} : (PQ)^* P = P(QP)^*$

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

* $I_{12} : (P+Q)R = PR + QR$ & $R(P+Q) = RP + RQ$

* If 2 RE P and Q are equivalent, we write $P = Q$. if, P and Q represents the same set of strings, we give the identities for RE which are useful for simplifying RE

Q. Write a RE for set of all languages whose(a) length is exactly 2 over alphabet (a, b)

$\Rightarrow \Sigma = \{a, b\}$

condn: length is exactly 2.

$\Rightarrow L = \{aa, ab, ba, bb\}$

$= aa + ab + ba + bb$

$= a(a+b) + b(a+b) = (a+b)(a+b)$

b) exactly 3

$\Rightarrow \Sigma = \{a, b\}$

$\Rightarrow L = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$

$= aaa + aab + aba + abb + baa + bab + bba + bbb$

$= aa(a+b) + ab(a+b) + ab(a+b) + bb(a+b)$

$= (a+b)(aa + ab + ab + bb)$

$= (a+b)(a(a+b) + b(a+b))$

$= (a+b)(a+b)(a+b)$

Q. Find RE for set of all string whose length is atmost 2.

$\Rightarrow \Sigma = \{a, b\}$

cond: Length is atmost 2

$L = \{\epsilon, a, b, ab, ba, aa, bb\}$

$= \epsilon + a + b + ab + ba + aa + bb$

$= \epsilon + a(b+a) + b(a+b) + a + b$

$= \epsilon + a + b + (a+b)(a+b)$

$= \epsilon + (a+b)(1 + (a+b))$

Q. Find RE for all strings of a) even length over (a,b)

$L = \{\epsilon, aa, bb, \cancel{aabb}, \cancel{aaaabbbb} \dots\}$

$\cancel{(aabb)^*}$ $\cancel{(ab)^*}$

$L = \{\epsilon, aa, ab, ba, bb, aaaa, abba, aabb \dots\}$

$= ((a+b)(a+b))^*$

b) odd length.

$L = \{\epsilon, a, b, aba, baa, \dots\}$

$= (a+b)((a+b)(a+b))^*$

c) No. of a's exactly 2. with any no. of b's

$L = \{aa\}$ $= b^*ab^*ab^*$

$= aa.$

Regular expression to Finite automata :

Prerequisits:

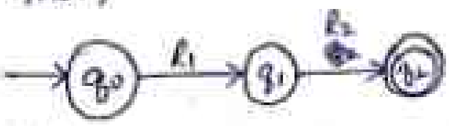
1) ϕ 4) $R_1 R_2$
2) ϵ 5) $R_1 + R_2^*$
3) R_1 6) R_1^*

1) for ϕ the initial state and final state is same

2) For ϵ we will have 2 states with ϵ

3) For a simple regular expression R1,

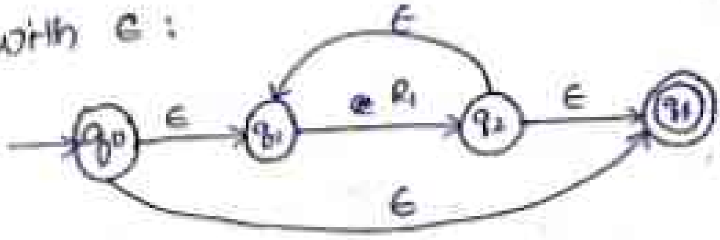
4) For 2 regular expressions which are represented as $R_1 R_2$ then



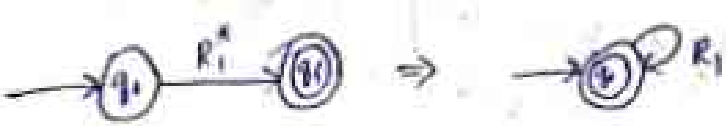
5) For concatination of 2 regular expressions $R_1 + R_2$,

6) Finite automata diagram for R_1^* is

a) with ϵ :



b) without ϵ :



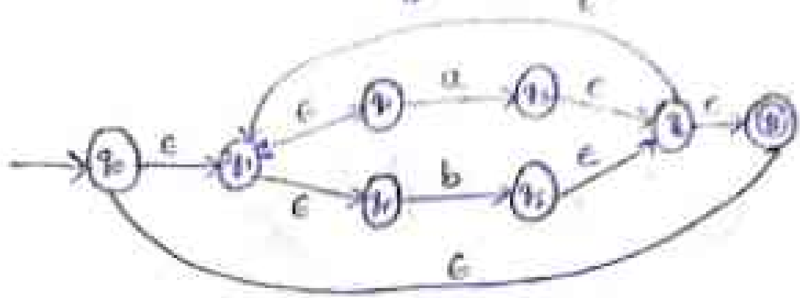
Q. $(a+b)^*$

Let $(a+b)^*$ be R_1^*

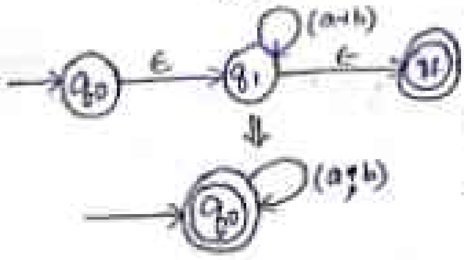
⇒ Finite automata for R_1^* is:

⇓

with ∈ ←



(or)

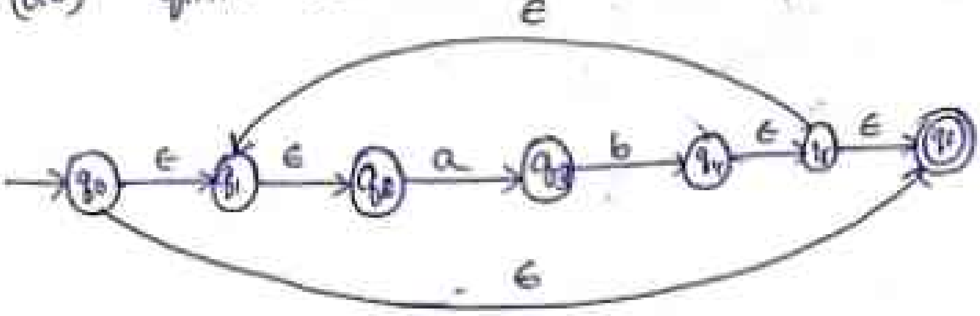


⇒ without ∈

Q) $(ab)^*$

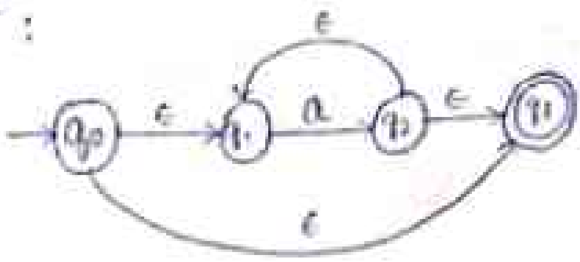
⇒ ab's finite automata is:

⇒ $(ab)^*$ finite automata is

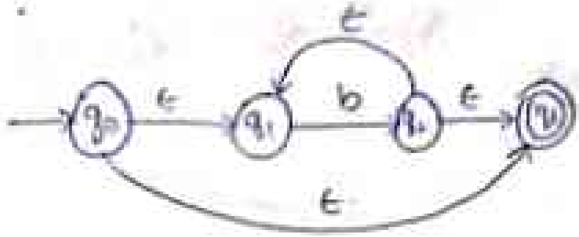


Q. $a^* + b^* + c^*$

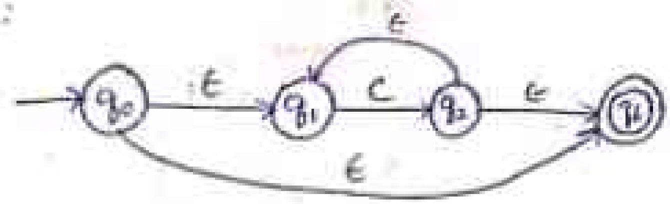
a^* :



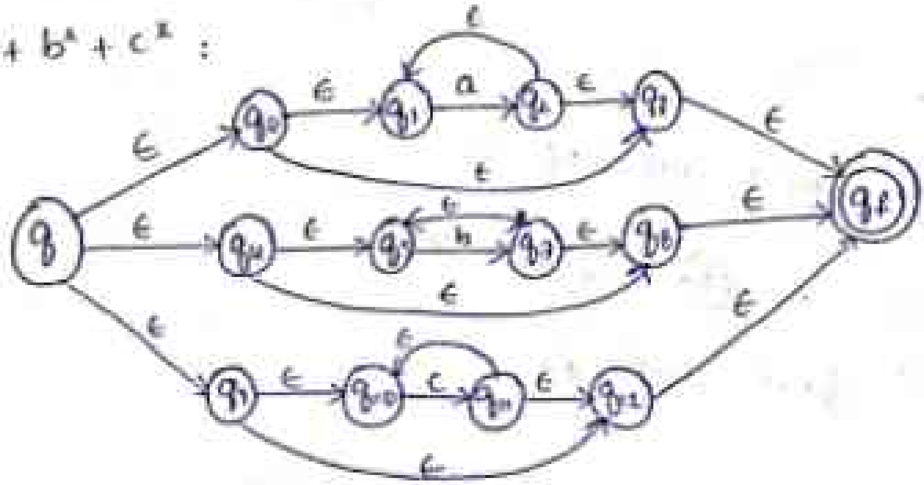
b^* :



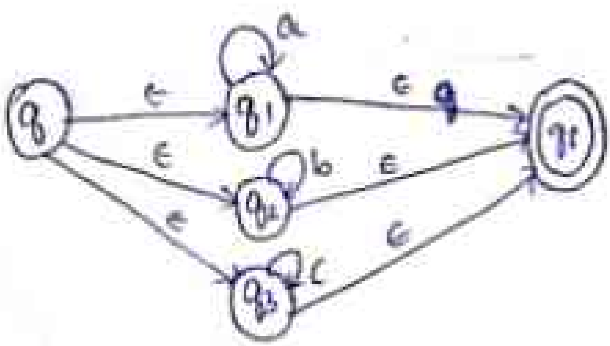
c^* :



→) $a^* + b^* + c^*$:

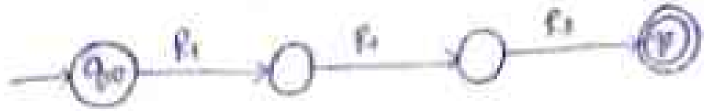


(or)



Q. Construct a finite automata equivalent to regular expression a) $\underset{R_1}{(a+b)^*}\ \underset{R_2}{aa}\ \underset{R_3}{(a+b)^*}$

$\Rightarrow$ $R_1 . R_2 R_3$:



$\Rightarrow$ 

$\Rightarrow$ Finite automata for $(a+b)^*$ would be

$\Rightarrow$ 

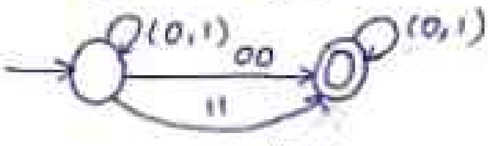
Finite automata for $R_1 R_2$ would be

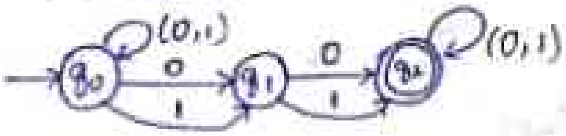
$\Rightarrow$ 

b) $\underset{R_1}{(0+1)^*}\ \underset{R_2+R_3}{(00+11)}\ \underset{R_4}{(0+1)^*}$

$\Rightarrow$ 

$\Rightarrow$ 

$\Rightarrow$ 

$\Rightarrow$ 

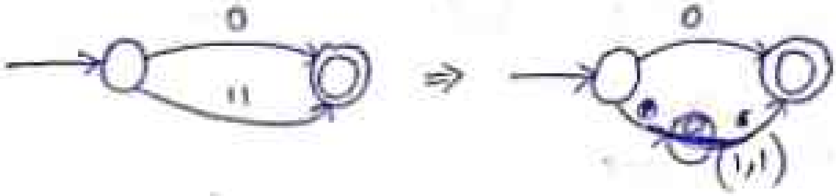
Q. Construct a finite automata with reduced states equivalent to regular expression

$$\underbrace{10}_{R_1} + \underbrace{(0+11)}_{R_2+R_3}\underbrace{0^*1}_{R_4}$$

$R_1 : 10$

$\Rightarrow$ 

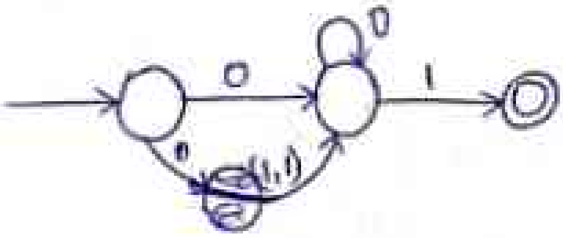
$R_2 + R_3 : 0 + 11$

$\Rightarrow$ 

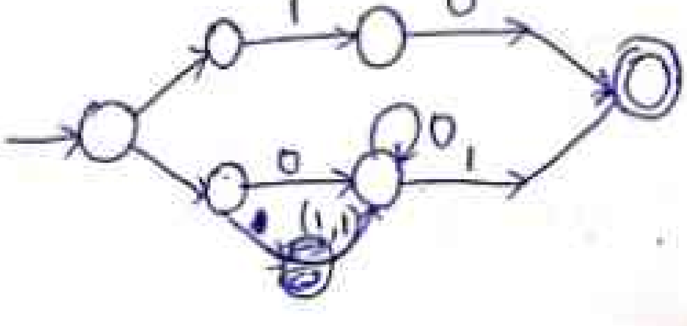
$R_4 : 0^* 1$

$\Rightarrow$ 

$\Rightarrow (R_2 + R_3) R_4$.

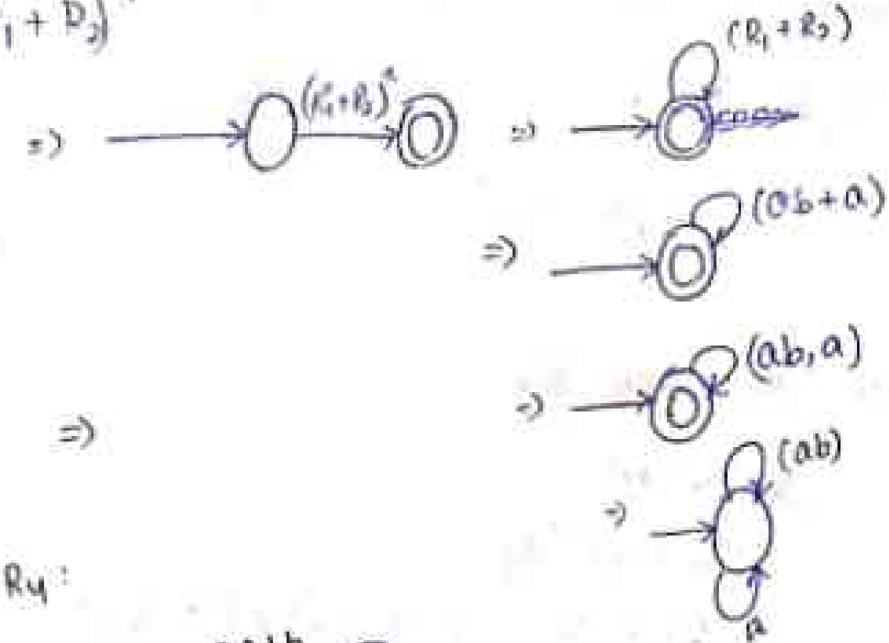
$\Rightarrow$ 

$R_1 + (R_2 + R_3) R_4$

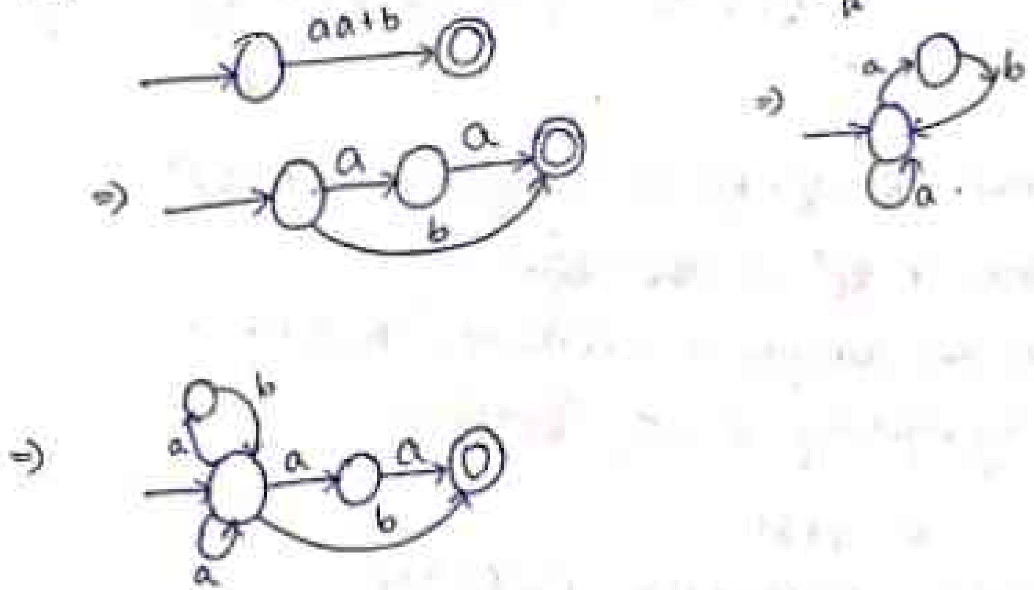


Q. $\underbrace{(ab+a)^*}_{R_1+R_2} \underbrace{(aa+b)}_{R_3+R_4}$

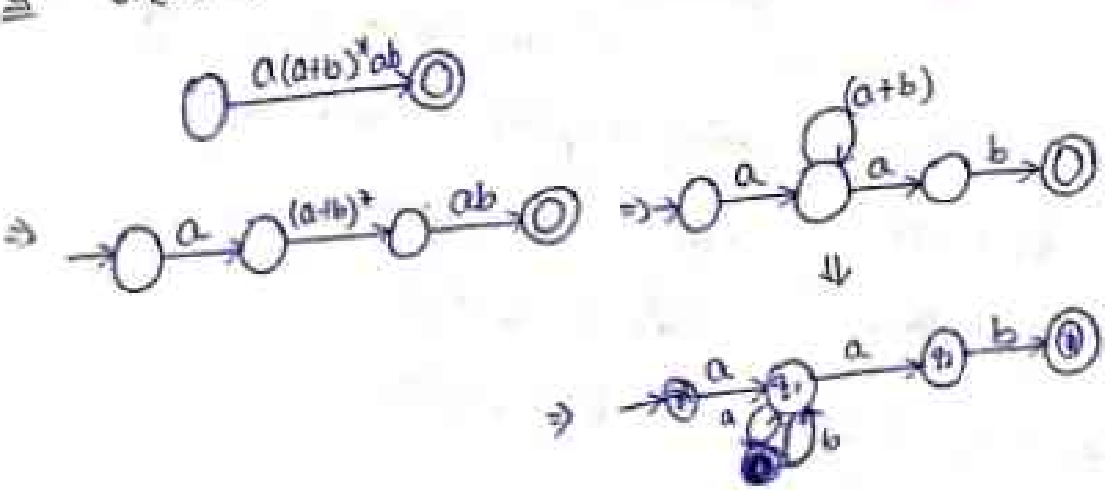
$\Rightarrow (R_1+R_2)^*$



$R_3 + R_4$:



Q. $a(a+b)^*ab$



Arden's theorem:

Let P & Q to be 2 RE over alphabet (Σ). If P doesnot contain ϵ then the following eqn in R

$$\Rightarrow \boxed{R = Q + RP} \quad , \{P \notin \epsilon , Q \in \epsilon\}$$

has a unique soln. $\boxed{R = QP^*}$

Proof:

equate both the R's

$$R = Q + RP$$
$$\Rightarrow Q + (QP^*) P$$
$$\Rightarrow Q + QP^*P = Q(\epsilon + P^*P) = QP^*$$
$$= R$$

Hence $R = Q + RP$ is Satisfied when $R = QP^*$

This means $R = QP^*$ is the soln.

To prove the uniqueness, consider $R = Q + RP$.

Here by replacing R with QP^*

$$R = Q + RP$$
$$= Q + (Q + RP)P = Q + QP + RP^2$$
$$= Q + QP + (Q + RP)P^2$$
$$= Q + QP + QP^2 + \cdots + QP^n + RP^{n+1}$$

Now replace R with QP^*

$$= Q + QP + QP^2 + \cdots + QP^n + (QP^*)P^{n+1}$$
$$= Q + QP + QP^2 + \cdots + QP^n + QP^*P^{n+1}$$
$$= Q(\epsilon + P + P^2 + P^3 + \cdots + P^n + P^*P^{n+1})$$
$$= QP^*$$
$$\Rightarrow \boxed{R = QP^*}$$

prove that LHS = RHS for given expression

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

LHS:

$$\Rightarrow (1+00^*1)(\epsilon + (0+10^*1)^*(0+10^*1))$$

Iq: $\epsilon + R^*R = R^*$

$$\Rightarrow (1+00^*1)((0+10^*1)^*)$$

$$\Rightarrow (\epsilon+00^*)1(0+10^*1)^* \Rightarrow 0^*1(0+10^*1)^*$$

$$= RHS.$$

Algebraic Method using Arden's theorem:

The following method is the extension of Arden's theorem. This is used to find RE recognised by a transition system.

The following assumptions are made:

The transition graph does not contain e-moves (ϵ)

→ It has only one initial state say V_1

→ Its vertices are $V_1, V_2, \ldots V_n$

→ V_i is the RE representing the set of strings accepted by the system eventhough V_i is the final state

→ α_{ij} denotes RE representing the set of labels of edges from V_i to V_j. When there is no such edge,

$\alpha_{ij} = \phi$

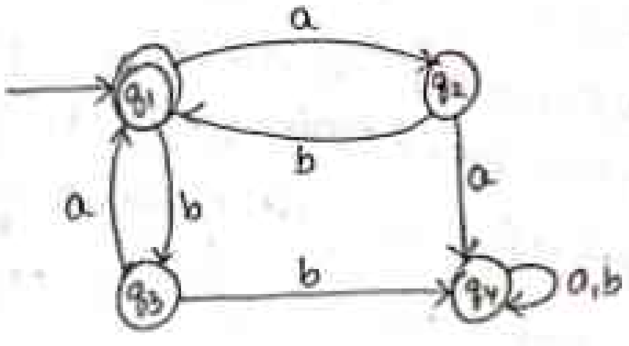
→ consequently we can get the following set of eqn's in $V_1, V_2 \ldots V_n$

$$V_1 = V_1\alpha_{11} + V_2\alpha_{21} + \dots + V_n\alpha_{n1}$$
$$V_2 = V_2\alpha_{12} + V_2\alpha_{22} + \dots\ V_n\alpha_{n2}$$
$$V_3 = V_3\alpha_{13} + V_3\alpha_{23} + \dots + V_n\alpha_{n3}$$
$$\vdots$$
$$V_n = V_n\alpha_{1n} + V_n\alpha_{2n} + \dots + V_n\alpha_{nn}.$$

By repeatedly applying substitutions and theorem of Arden, we can express V_i interms of α_{ij}'s.

For getting set of strings, recognised by transition systems, we have to take the union of all V_i's corresponding to final states.

Q. prove that the finite automata whose transition diagram is given, accepts the set of all strings over alphabet $(\Sigma) = \{a, b\}$ with equal no. of a's & b's such that each prefix has atmost 1 more a than b. and atmost one more b than a's.



$$\Rightarrow \quad q_1 = q_2 b + q_3 a + \epsilon \quad \text{——} ①$$
$$q_2 = q_1 a \quad \text{——} ②$$
$$q_3 = q_1 b \quad \text{——} ③$$
$$q_4 = q_2 a + q_3 b + q_4 a + q_4 b \quad \text{——} ④$$

Substitute q_2, q_3 in ①

$\Rightarrow q_1 = (q_1 a)b + (q_1 b)a + \epsilon$

$= q_1 ab + q_1 ba + \epsilon$

$= q_1(ab+ba) + \epsilon$

$q_1 = \epsilon + q_1(ab+ba) \longrightarrow$ is in the form $R = Q + RP$

$\Rightarrow R = QP^*$

$q_1 = \epsilon \cdot (ab+ba)^*$

$q_1 = \underline{(ab+ba)^*}$

Q. Describe in English the set accepted by FA whose Transition diagram is

$\Rightarrow q_1 = 0q_1 + \epsilon$ —①

$q_2 = 1q_1 + q_2 1$ —②

$q_3 = 0q_2 + q_3(0+1)$ —③

eqn ① is in form $R = Q + RP$

$\Rightarrow R = QP^*$

$\Rightarrow q_1 = \epsilon 0^*$

$\Rightarrow \boxed{q_1 = 0^*}$

Solve eqn ② $\Rightarrow q_2 = 0^*1 + q_2 1$

~~$q_2 = q_2 1 + 0^*1$~~

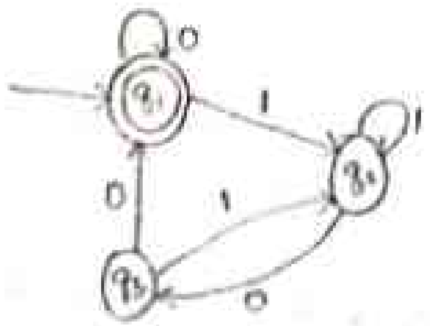
$\Rightarrow \boxed{q_2 = 0^*11^*}$

$\Rightarrow$ Final state q_1 and q_2 :

$q_1 + q_2 = 0^* + 0^*11^* = 0^*(\epsilon + 11^*)$

$\underline{q_1 + q_2 = 0^*1^*}$ $\left[\begin{array}{l} \text{by } I_9 \\ = \epsilon + RR^* \\ = R^* \end{array}\right]$

Q.



$q_1 = 0q_1 + 0q_3 + \epsilon$ — (1)

$q_2 = 1q_2 + 1q_1 + 1q_3$ — (2)

$R = Q + RP$

$q_3 = 0q_2$ — (3)

Substitute (3) in (2)

$\Rightarrow \cancel{q_1 = 0q_1 + (q_3 + \epsilon)}$

$\cancel{R}$

$q_2 = 1q_1 + 1q_2 + q_2 01$

$q_2 = 1q_1 + q_2(1+01)$

$R = Q + R\ P$

$\Rightarrow R = QP^*$

$q_2 = 1q_1(1+01)^*$ — (4)

$\Rightarrow q_1 = 0q_1 + 00q_2 + \epsilon$

$= 0q_1 + 00(1q_1(1+01)^*) + \epsilon$

$= q_1(0 + 1(1+01)^* 00) + \epsilon$

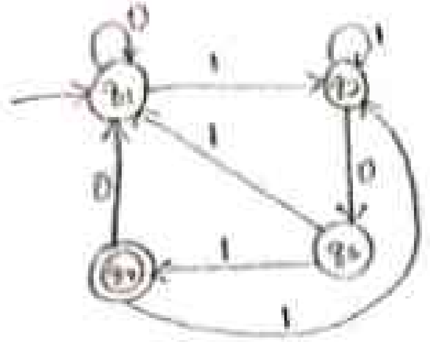
$\Rightarrow q_1 = \epsilon + q_1(0 + 1(1+01)^* 00)$

$R = Q + R\ P$

$\Rightarrow R = QP^*$

$\Rightarrow q_1 = \epsilon(0 + 1(1+01)^* 00)^*$

Q.



$\Rightarrow q_1 = 0q_1 + 1q_3 + 0q_4 + \epsilon$

$q_2 = 1q_1 + 1q_2 + 1q_4$

$q_3 = 0q_2$

$q_4 = 1q_3$

$\longrightarrow$ Substitute all req. values in q_2.

$\Rightarrow q_2 = 1q_1 + 1q_2 + 1q_4$

$= 1q_1 + 1q_2 + 110q_2$

$q_2 = 1q_1 + q_2(1+110)$

$R = Q + RP$

$\Rightarrow R = QP^* \Rightarrow q_2 = 1q_1(1+110)^*$ — ①

$\Rightarrow q_3 = 01q_1(1+110)^*$ — ②

$\Rightarrow q_4 = 101q_1(1+110)^*$ — ③

Substitute ①,②,③ in q_1

$\Rightarrow q_1 = 0q_1 + 1q_3 + 01q_3 + \epsilon$

$= 0q_1 + 101q_1(1+110)^* + 0101q_1(1+110)^* + \epsilon$

$q_1 = q_1(0 + 101(1+110)^* + 0101(1+110)^*) + \epsilon$

$R = R \quad P \quad + Q$

$\Rightarrow R = QP^* \Rightarrow q_1 = \epsilon(0 + (1+110)^*(101+0101))^*$ — ④

$\Rightarrow$ Substitute ④ in ③

$\Rightarrow q_4 = 101\,\epsilon(0 + (1+110)^*(101+0101))^*(1+110)^*$

Q.

$\Rightarrow \quad q_0 = aq_0 + bq_0 + \epsilon$

$\Rightarrow q_0 = \epsilon + q_0(a+b)$

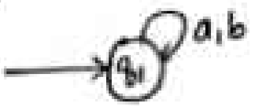
$R = Q + RP$

$\Rightarrow R = QP^*$

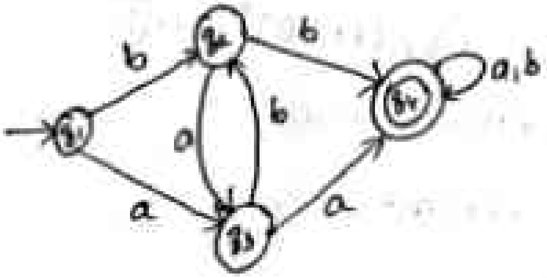
$\Rightarrow q_0 = \epsilon(a+b)^*$

$q_0 = (a+b)^*$

Q. Find the set of strings over $\Sigma = \{a, b\}$ recognised T.S

a) 

b) 

c) 

a) $q_1 = aq_1 + bq_1 + \epsilon$

$\because$ No final state $\Rightarrow$ Regular expression is ϕ

b) $q_1 = aq_1 + bq_1 + \epsilon$

$= q_1(a+b) + \epsilon$

$\Rightarrow q_1 = \epsilon(a+b)^*$

$q_1 = (a+b)^*$

c) $q_1 = \epsilon$

$q_2 = bq_3 + bq_1 = b(q_3 + \epsilon) = bq_3$ $\boxed{\because \epsilon + x = x}$

$q_3 = aq_2 + aq_1 = abq_3 + a\epsilon = a(bq_3 + \epsilon)$

$= abq_3$

$q_4 = bq_2 + aq_3 + q_4(a+b)$

$\Rightarrow$

$q_4 = bq_3 + aabq_3 + q_4(a+b)$

$q_4 = q_3(b + aab) + q_4(a+b)$

$\Rightarrow q_4 = bbq_3 + aq_3 + q_4(a+b)$

$\Rightarrow q_2 = bq_3 + b\epsilon$