Introduction to Automata Theory, Formal Languages and Computation

Shyamalendu Kandar

Assistant Professor, Computer Science and Engineering

Haldia Institute of Technology

Haldia, West Bengal

225 to 228

2020 December 28

b

Solution:

(i) The strings that we get from the RE 10*1 are {11, 101, 1001, 10001…..}. From the set, we can conclude that the strings that are generated from the REs are beginning with 1 and ending with 1. In between 1 and 1, there are any numbers of 0.

So, we can describe in English like this:

Set of all strings beginning and ending with 1 with any number of 0 in between the two 1s.

(ii) The strings that we get from the RE a*b* are a* . b*, i.e.,

{Λ, a, aa, aaa, ……}.{Λ, b, bb, bbb, …….}   {Λ, a, b, aa, bb, ab, abb, aab, ……..}.

From the set, we can conclude that the strings that are generated from the RE contain any number of 'a' followed by any number of 'b'.

So, we can describe in English like this:

Set of all strings of 'a' and 'b' containing any number of 'a' followed by any number of 'b'.

(iii) The strings that we get from the RE (ab)* are {Λ, ab, abab, ababab, ……….}.

In English, it is described as:

Set of all strings of a and b with equal number of a and b containing 'ab' as repetition.

(iv) (a*   b*) means any combination of a or any combination of b. c* means any combination of c, which is concatenated with (a*   b*).

In English, it can be described as:

Any combination of 'a' or any combination of 'b' followed by any combination of 'c'.

(v) The strings that we get from the RE (00)*(11)*1 are {1, 001, 111, 00111, ……}. In the string, there are always odd number of 1 and even number of 0.

In English, it can be described as:

Set of all strings of '0' and '1' with even number of 0 followed by odd number of 1.

Alternatively, the RE described in English language can be denoted using alphabet and symbols. The following examples describe this.

Example 5.2. Built regular expression of the following
(i) An RE consists of any combination of a and b, beginning with a and ending with b.
(ii) A language of any combination of a and b containing abb as a substring.
(iii) RE of a and b containing at least 2 'a's.
(iv) Write an RE for the language $L = \{a^n b^m | where\, m + n\, is\, seven\}$.
(v) An RE of a and b, having exactly one 'a'.

Solution:

(i) The RE consists of any combination of a and b, i.e., (a + b)*. But the RE starts with a and ends with b. In between a and b, any combination of a and b occurs. So, the RE is

$$L = a(a + b)*b.$$

(ii) In the RE, abb occurs as a substring. The RE consists of any combination of a and b. Before the substring, abb may occur at the beginning, at the middle, or at the last. If abb occurs at the beginning, then after abb there is any combination of a and b. If abb occurs at the middle, then before and after abb there are any combination of a and b. If abb occurs at the last, then before abb there is any combination of a and b.

The RE is L = (a + b)*abb(a + b)*.

(iii) In the RE, there are at least two 'a's. The expression consists of a and b. Therefore, any number of 'b' can occur before the first 'a' and before the second 'a', i.e., after the first 'a' and after the second 'a'. So, the RE will be L = b*ab*ab*.

(iv) The RE consists of n number of 'a' followed by m number of 'b'. But m + n is always even. This is possible if

(a) m and n both are even or (b) m and n both are odd.

If m and n are both even, then the expression is (aa)*(bb)*.

If m and n are both odd, then the expression is (aa)*a(bb)*b.

By combining these two, the final RE becomes

$$L = \ (aa)^*(bb)^* + \ (aa)^*a(bb)^*b.$$

(v) In the RE, there is exactly one 'a'. Before and after a, there is any combination of b.

Therefore, the RE is L = b*ab*.

5.3   Identities of Regular Expression

An *identity* is a relation which is tautologically true. In mathematics, an equation which is true for every value of the variable is called an identity equation. As an example, $(a+b)^2 = a^2 + 2ab + b2$ is an identity equation. Based on these identities, some other problems can be proved. In the RE also, there are some identities which are true for every RE. In this section, we shall discuss those identities related to RE.

```
1. ∅ + R = R + ∅ = R
```

```
Proof: LHS : ∅ + R
```

$$= \emptyset \cup R$$
$$= \{\} \cup \{Elements of R\} = R = RHS$$

```
2. ∅R = R∅ = ∅
   Proof: LHS : ∅R
```
$$= \emptyset \cap R$$
$$= \{\} \cap \{Elements of R\} = \emptyset = RHS$$

```
Note: In both the previous cases, ∅ denotes a null set, which contains nothing.
3. ΛR = RΛ = R
```

```
Proof: LHS : ΛR
       = Null string concatenated with any symbol of R
       = Same symbol∈ R = R = RHS
4. Λ* = Λ & ∅* = Λ
```

```
Proof: LHS : Λ* = {Λ, ΛΛ, ΛΛΛ.}
            = {Λ, Λ, Λ, ..}[according to identity(3)]
            = Λ = RHS.
```

```
Same for ∅*.
```

5. $R + R = R$

   `Proof:` $LHS : R + R = R(\Lambda + \Lambda) = R$
   $= R$ `[according to identity (3)]` $= RHS$

6. $R^* R^* = R^*$

   `Proof:` $LHS : R^* R^* = \{\Lambda, R, RR\}\{\Lambda, R, RR\}$
   $= \{\Lambda\Lambda, \Lambda R, \Lambda RR, .., R\Lambda, RR, , RR, RRR, .\}$
   $= \{, R, RR, RRR, ..\}[usingidentity(3)] = R* = RHS$

7. $R^* R = RR^*$

   `Proof:` $LHS : R^* R = \{\Lambda, R, RR, RRR, ..\}R$
   $= \{\Lambda R, RR, RRR, RRRR, ..\}$
   $= R\{\Lambda, R, RR, RRR, ..\} = RR^* = RHS$

8. $(R^*)^* = R^*$

   `Proof:` $LHS : (R^*)^* = \{\Lambda, R^* R^*, R^* R^* R^*, ..\}$
   $= \{\Lambda, R^*, R^*, ..\}$ `[using identity (6)]`
   $= \{\Lambda, \{\Lambda, R, RR, RRR, ..\}, \{\Lambda, R, RR, RRR, ..\}, .\}$
   $= R^* = RHS$

9. $\Lambda + RR^* = \Lambda + R^* R = R^*$

   `Proof:` $LHS : \Lambda + RR^* = \Lambda + R\{\Lambda, R, RR, RRR, ..\}$
   $= \Lambda + \{R\Lambda, RR, RRR, RRRR, ..\}$
   $= \Lambda + \{R, RR, RRR, RRRR, ..\}$
   $= \{\Lambda, R, RR, RRR, RRRR, ..\} = R^* = RHS$

10. $(PQ)^* P = P(QP)^*$

    `Proof:` $LHS : (PQ)^* P = \{\Lambda, PQ, PQPQ, PQPQPQ, ..\}P4=$ `\{P,PQP, PQPQP, PQPQPQP…,..\}4`
    $= P\{\Lambda, QP, QPQP, QPQPQP, .\}$
    $= P(QP)^* = RHS$

11. $(P + Q)^* = (P^* Q^*)^* = (P^* + Q^*)^*$ (`''DMorgans theorem`)
12. $(P + Q)R = PR + QR$

    `Proof:` $Let a \in (P + Q)R$
    $= a \in PR or QR = RHS.$

Example 5.3. From the identities of RE, prove that

$$(1 + 100^*) + (1 + 100^*)(0 + 10^*)(0 + 10^*)^* = 10^*(010^*)^*.$$

Solution:LHS.

$$(1 + 100^*) + (1 + 100^*)(0 + 10^*)(0 + 10^*)^*$$
$$= (1 + 100^*)(\Lambda + (0 + 10^*)(0 + 10^*)^*)$$
$$= (1 + 100^*)(0 + 10^*)^* \text{ (according to } \Lambda + RR^* = R^*)$$
$$= 1(\Lambda + 00^*)(0 + 10^*)^*$$
$$= 10^*(0 + 10^*)^* = RHS.$$

Example 5.4. From the identities of RE, prove that the following three are equivalent

(i) $(011((11)^* + (01)^*)^*)^*011$

(ii) $011(((1+0)1)^*011)^*$

(iii) $011(((11)^*(01)^*)^*011)^*$

Solution:Let$P = (11) and Q = (01).$

```
We know(P + Q)* = (P*Q*)* = (P* + Q*)*.
Thus, ((11) + (01))* = ((11)*(01)*)*[inc] = ((11)* + (01)*)*[ina] = ((1+0)1)*[inb].
 Consider 011 as R and (P + Q)* as S.
We know that   (RS)*R = R(SR)*
Thus, (011((11)* + (01)*)*)*011 = 011((11)* + (01)*)*011)*
```
$$= 011((11) + (01))^*011)^*$$
$$= 011(((1+0)1)^*011)^*.$$
```
Similarly, (011((11) * +(01)*)*)*011 = 011((11)* + (01)*)*011)*
```
$$= 011((11) + (01))^*011)^*$$
$$= 011(((11)^*(01)^*)^*011)^*.$$
```
As  (i) ≡ (ii) and (i) ≡ (iii), (i) ≡ (ii) ≡ (iii).
```

Example 5.5. From the identities of RE, prove that

$$10 + (1010)^*[\wedge + (1010)6^*] = 10 + (1010)^*.$$

Solution: LHS $10 + (1010)^*[\wedge + (1010)^*]$

$$\Rightarrow 10 + \wedge(1010)^* + (1010)^*(1010)^*$$
$$\Rightarrow 10 + (1010)^*(1010)^* As \wedge R = R and RR = R$$
$$\Rightarrow 10 + (1010)^* = RHS. As R + R = R.$$

5.4   The Arden's Theorem

The Arden's theorem is used to construct the RE from a transitional system of FA.

Theorem 5.1.   Statement: Let P and Q be two REs over Σ. If P does not contain Λ, then the equation $R = Q + RP$ has a unique (one and only one) solution, $R = QP^*$.

   Proof: Now, point out the statements in the Arden's theorem in general form.

☐ **P** and **Q** are two REs.

☐ **P** does not contain the **Λ** symbol.

☐ **R = Q + RP** has a solution, i.e., **R = QP***.

☐ This solution is the one and only solution of the equation.

   If R = QP* is a solution of the equation R + Q = RP, then by putting the value of R in the equation, we shall get the value '0'.

$$R = Q + RP$$
$$R - Q - RP = 0$$
$$LHS R - Q - RP$$