Finite representation of Languages.

Languages can be represented in close forms using the notations $\cdot, \cup, *$

<u>Example</u>

Let $L = \{ w \in \{0,1\}^* \mid$ w has two or three occurrance of 1, the first and the second of which are not consecutive $\}$.

L can be represented as follows

$$L = 0^* 1 0^+ 1 0^* 1 0^* + 0^* 1 0^+ 1 0^*$$

$$= 0^* 1 0^+ 1 (0^* + 1 0^*) \text{ //}$$

The closed form of the above types are called regular expression of a given language.

Formally, a regular expression over an alphabet Σ is defined as follows-

(1) ϕ and each symbol of Σ are regular expressions

(2) if α, β are regular expressions then so is $\alpha\beta$

(3) if α, β are regular expression then so is $\alpha + \beta$

(4) if α is a regular expression then so is α^*

The language represented by a regular expression is called a regular language.

The language corresponding to a regular expression is defined as follows-

ϕ

(1) $L(\phi) = \phi$, $L(a) = \{a\}$ for $\forall\, a \in \Sigma$

(2) If α, β are regular expressions

then $L(\alpha\beta) = L(\alpha)\,L(\beta)$

(3) if α, β are regular expression

the $L(\alpha + \beta) = L(\alpha) + L(\beta)$

(4) if α is a regular expression

then $L(\alpha^*) = (L(\alpha))^*$

Problem

Find a regular expression of the set of binary strings which have at least one occurrence of the substring 001.

Ans Such strings can be written as $x001y$, where x, y could be any string.

$\therefore$ The r.e of the language is $(0+1)^* 001 (0+1)^*$

Problem

Find a regular expression for the set of binary string with the property that none of its prefixes has two more 0's than 1's nor two more 1's than 0's.

Solu^n

Suppose S be a string in L of length n.

$S = x_1 x_2 \ldots x_n$

claim 1 $\quad x_1 \neq x_2$

claim 2 $\quad x_3 \neq x_4$

claim 3 $\quad x_5 \neq x_6$

$\vdots$

$x_{2j-1} \neq x_{2j}$ for $j \leq \lfloor \frac{n}{2} \rfloor$

From the above observation, the r.e of L is

$$(01+10)^*(0+1+\epsilon)$$