Lecture Three

Regular Expression (RE)

Regular Expression is a set of symbols, thus if alphabet= {a, b}, then aab, a, baba, bbbbb, and baaaaa would all be strings of symbols of alphabet.
In addition, we include an empty string denoted by $\wedge$ which has no symbols in it. We now introduce the use of the Kleene star applied not to a set but directly to the letter x and written as (x*). The simple expression x* will be used to indicate some sequence of x's.

$x^* = \{ \wedge, x, x^2, x^3, x^4, ... \}$ $= \{x^n \text{ for } n = 0, 1, 2, 3, 4, ...\}$

Examples:

1- (ab)* = {∧, ab, abab, ababab, … }

2- ab*a = {aa, aba, abba, abbba, abbbba, …}

3- a*b* = {∧, a, b, aa, ab, bb, aaa, aab, abb, bbb, aaaa, …}

✍Notice that **ba** and **aba** are not in this language. Also we should be very careful to observe that **a*b* ≠ (ab)***

4- $L_1 = \{x^{odd}\}$ = x(xx)* **or** (xx)*x = {x, xxx, xxxxx, …}

5- $L_2 = \{x^{even}\}$ = xx(xx)* **or** (xx)*xx **or** (xx)* = {∧, xx, xxxx, …}

6- Consider the language L_3 defined over the alphabet ∑ = {a, b, c}, All the words in L_3 begin with an **a** or **c** and then are followed by some number of **b's**. We may write this as:

$$L_3 = (a + c)b^*$$

7- Consider a finite language L_4 that contains all the strings of **a's** and **b's** of length exactly three.

$$L_4 = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$$

So we may write:

$L_4 = (a + b)(a + b)(a + b)$ **or** $(a + b)^3$

In general, if we want to refer to the set of all possible strings of **a's** and **b's** of any length, we could write:

$$L_4 = (a + b)^*$$

8- Construct RE for all words that begin with the letter **a**: a(a + b)*

9- All words that begin with an **a** and end with **b** can be defined by the expression: a(a + b)*b

10- The language of all words that have at least two **a's** can be described by the expression:

(a + b)*a(a + b)*a(a + b)* **<u>or</u>** b*ab*a(a + b)*

11- The language of all words that have at least one **a** and at least one **b**:

(a + b)*a(a + b)* b(a + b)* **<u>or</u>** bb*aa*

12- The words of the form some **b's** followed by some **a's**. These exceptions are all defined by the regular expression: bb*aa* $\equiv b^+a^+$ ***<u>Homework:</u>***

1- Find a regular expression over the alphabet {a, b}:

a. L_1 = {All strings that contain exactly three a's}
b. L_2 = {All strings that end with ab}
c. L_3 = {All strings in which letter a is even number}
d. L_4 = {All strings that contain exactly two successive a's.}

2- Find the output (words) for the following regular expressions:

a. aa* b
b. bba*a
c. (a + b)* ba
d. (0+1)* 00 (0+1)*
e. (11 + 0)* (0+11)*
f. 01* + (00+101)*
g. (a+b)* abb$^+$
h. (((01+10)* 11)* 00)*