COMSCI 33 (AUTOMATA THEORY)

Lesson 2: Regular Expression

Regular Language

- In theoretical computer science and formal language theory, a **regular language** is a formal language that can be expressed using a regular expression

Regular Expression

- A language-defining notation
- A declarative way to express the string we want to accept
- The Regular Expressions (RE) over an alphabet sigma are the strings over the alphabet sigma(Σ) union with parentheses, star operator and plus sign.

  r = Σ ∈ { (,), *, + }

Definition of a Regular Expression

- Let Σ be an alphabet. The regular expressions over Σ are:
  - Ø Represents the empty set { }
  - ε Represents the set {ε}
  - a Represents the set {a}, for any symbol a in Σ

  Let r and s be regular expressions that represent the sets R and S, respectively.
  - r+s Represents the set R U S
  - rs Represents the set RS
  - r^* Represents the set R*
  - (r) Represents the set R
- If r is a regular expression, then L(r) is used to denote the corresponding language.

Example: Let Σ = {0, 1}

(0 + 1)*	All strings of 0's and 1's
01*	0 followed by any number 1's
0(0 + 1)*	All strings of 0's and 1's, beginning with a 0
(0 + 1)*1	All strings of 0's and 1's, ending with a 1

Practice Problems

- What is the RE of the following, Let Σ = {a, b}
  - L = {b, ba, baa, baaa, …}
  - L = { aa, bb, ab, ba}
  - L = {a, aaa, aaaaa, aaaaaaa, aaaaaaaaa, ….}

  - L = { w∈{a,b}* / w starts with a}
  - L = { w∈{a,b}* / w contains the substring aa}
  - L = { w ∈ Σ* / the third to the last symbol in w is an a
  - L = { w ∈ Σ* / |w| < 2 }
  - L = { w ∈ Σ* / w contains exactly 3 a's}

Assignment

- L = { w∈{a,b}* / w contains aa or bb}
- L = { w∈{a,b}* / w contains the substring aba}
- L = { w∈{a,b}* / w starts and ends in b}
- L = { w∈{a,b}* / |w|=2k where k=0,1,2,3,…}
- L = { w∈{a,b}* / w ends in ab}
- L = { w∈{a,b}* / w contains at most 1 b}
- L = { w∈{a,b}* / w contains the substring aa}
- L = { w∈{a,b}* / w contains 0 or more clumps of aba's}
- L = { w∈{a,b}* / second symbol in w is an a}
- L = { w∈{a,b}* / |w| = 3k + 1 where k=0,1,2,3,…}
- L = { w∈{a,b}* / w contains all a's or all b's}
-