THEORY OF AUTOMATA

&

FORMAL LANGUAGES

HANDOUTS 04

Department of Information Technology
Hazara University, Mansehra

Regular Expressions

As discussed earlier that a* generates Λ, a, aa, aaa, $\cdots$ and a+ generates a, aa, aaa, aaaa, $\cdots$, so the language L1 = {Λ, a, aa, aaa, $\cdots$} and L2 = {a, aa, aaa, aaaa, $\cdots$} can simply be expressed by a* and a^+, respectively.

a* and a^+ are called the regular expressions (RE) for L1 and L2 respectively.

Note a^+, aa* and a*a generate L2.

Recursive definition of Regular Expression(RE)

Step 1: Every letter of Σ including Λ is a regular expression.

Step 2: If r1 and r2 are regular expressions then

(r1)

r1 r2

r1 + r2 and

r1*

are also regular expressions.

Step 3: Nothing else is a regular expression.

Method 3 (Regular Expressions)

Consider the language L={Λ, x, xx, xxx,$\cdots$} of strings, defined over Σ = {x}.

We can write this language as the Kleene star closure of alphabet Σ or L=Σ*={x}* .

This language can also be expressed by the regular expression x*.

Similarly the language L={x, xx, xxx,$\cdots$}, defined over Σ = {x}, can be expressed by the regular expression x^+.

Now consider another language L, consisting of all possible strings, defined over Σ = {a, b}. This language can also be expressed by the regular expression (a + b)*.

Now consider another language L, of strings having exactly one a, defined over Σ = {a, b}, then its regular expression may be b*ab*.

Now consider another language L, of even length, defined over Σ = {a, b}, then its regular expression may be ((a+b)(a+b))*.

Now consider another language L, of odd length, defined over Σ = {a, b}, then its regular expression may be (a+b)((a+b)(a+b))* or ((a+b)(a+b))*(a+b).

Remark

It may be noted that a language may be expressed by more than one regular expression, while given a regular expression there exist a unique language generated by that regular expression.

Example

Consider the language, defined over

Σ = {a , b} of words having at least one a, may be expressed by a regular expression (a+b)*a(a+b)*.

Consider the language, defined over Σ = {a, b} of words having at least one a and one b, may be expressed by a regular expression

(a+b)*a(a+b)*b(a+b)*+ (a+b)*b(a+b)*a(a+b)*.

Consider the language, defined over Σ ={a, b}, of words starting with double a and ending in double b then its regular expression may be

aa(a+b)*bb

Consider the language, defined over Σ ={a, b} of words starting with a and ending in b OR

starting with b and ending in a, then its regular expression may be

a(a+b)*b+b(a+b)*a

An important example

The Language EVEN-EVEN

Language of strings, defined over Σ={a, b} having even number of a's and even number of b's. i.e.

EVEN-EVEN = {Λ, aa, bb, aaaa,aabb,abab, abba, baab, baba, bbaa, bbbb,$\cdots$},

its regular expression can be written as

(aa+bb+(ab+ba)(aa+bb)*(ab+ba))*

Note

It is important to be clear about the difference of the following regular expressions

r1 = a*+b*

r2 = (a+b)*

Here r1 does not generate any string of concatenation of a and b, while r2 generates such strings.

Equivalent Regular Expressions

Definition

Two regular expressions are said to be equivalent if they generate the same language.

Example

Consider the following regular expressions

r1 = (a + b)* (aa + bb)

r2 = (a + b)*aa + (a + b)*bb

then both regular expressions define the language of strings ending in aa or bb.

Note

If r1 = (aa + bb) and r2 = (a + b) then

r1+r2 = (aa + bb) + (a + b)

r1r2 = (aa + bb) (a + b)

= (aaa + aab + bba + bbb)

(r1)* = (aa + bb)*

Regular Languages

Definition

The language generated by any regular expression is called a **regular language**.

It is to be noted that if r1, r2 are regular expressions, corresponding to the languages L1 and L2

then the languages generated by r1+ r2, r1r2(or r2r1) and r1*(or r2*) are also regular

languages.

Note

It is to be noted that if L1 and L2 are expressed by r1 and r2, respectively then the language expressed by

r1+ r2, is the language L1 + L2 or L1 ∪ L2

r1r2, , is the language L1L2, of strings obtained by prefixing every string of L1 with every string of L2

r1*, is the language L1*, of strings obtained by concatenating the strings of L, including the null string.

Example

If r1 = (aa+bb) and r2 = (a+b) then the language of strings generated by r1+r2, is also a regular language,

expressed by (aa+bb) + (a+b)

If r1 = (aa+bb) and r2 = (a+b) then the language of strings generated by r1r2, is also a regular language, expressed by (aa+bb)(a+b)

If r = (aa+bb) then the language of strings generated by r*, is also a regular language, expressed by (aa+bb)*

All finite languages are regular

Example

Consider the language L, defined over Σ = {a,b}, of strings of length 2, starting with a, then

L = {aa, ab}, may be expressed by the regular expression aa+ab. Hence L, by definition, is a regular language.

Note

It may be noted that if a language contains even thousand words, its RE may be expressed, placing ' +' between all the words.

Here the special structure of RE is not important.

Consider the language L = {aaa, aab, aba, abb, baa, bab, bba, bbb}, that may be expressed by a RE

aaa+aab+aba+abb+baa+bab+bba+bbb, which is equivalent to (a+b)(a+b)(a+b).