Regular Expressions (Examples)

Course: Theory of Automata-II

Topic: Regular Expressions (Examples)

Instructor: Dr. Ahmed Mateen

Regular Expression

Write RE for the following languages for $\Sigma = \{a,b\}$

- The language of all words

  (a+b)*
- All words ending with **b**

  (a+b)*b
- All words that start with **a**

  a(a+b)*
- The language of all strings, not beginning with b

  a(a+b)*+ Λ
- All words that start with a double letter

  (aa+bb)(a+b)*
- All words that contain **at least one double letter**

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

Regular Expression cont…

- All words that contain **at least two a's or two b's**
  b*ab*ab* + a*ba*ba*
- All words that start and end with a double letter
  (aa+bb)(a+b)*(aa+bb)
- All words of length >=3
  (a+b)(a+b)(a+b)(a+b)*or (a+b)(a+b)(a+b)+
- All words that contain exactly one **a or exactly one b**
  b*ab* + a*ba*
- All words that don't end at **ba**
  (a+b)*(aa+ab+bb)
- All strings of a's and b's in which either the strings are all b's or else there is an a followed by some b's
  b*+ab*

Regular Expression cont…

- Language of all words that have at least two **a's**

  (a+b)* a (a+b)* a (a+b)*
- Language of all words that have **at least one a and at least one b**

  (a+b)* a (a+b)* b (a+b)* + (a+b)* b (a+b)* a (a+b)*
- Language of all words that have **at least one a or at least one b**

  (a+b)*a(a+b)* + (a+b)*b(a+b)*
- The languages L, of even length, defined over Σ = {a, b}

  ((a+b)(a+b))*
- The languages L, of odd length, defined over Σ = {a, b}

  ((a+b)(a+b))*(a+b)
- The strings of length 2, starting with **a**,

  aa+ab

Regular Expression cont…

- EVEN-EVEN (Σ = {a, b}) i.e. = {Λ, aa, bb, aaaa, aabb, abab, abba, baab, baba, bbaa, bbbb, …}

  RE sets:
- R1 = (aa+bb)*
- R2 = ((ab+ba)(ab+ba))*

  =>
- R.E. For EVEN-EVEN =
- (aa + bb + (ab + ba)(aa + bb)* (ab + ba))*

Regular Expression cont…

- ODD-ODD ($\Sigma = \{a, b\}$)

R.E : $(a+b)(aa+bb+(ab+ba)(aa+bb)^*(ab+ba))^*$

- // Think and Solve Your Own

Regular Expression cont…

- The set of all strings of a's and b's that have atleast two letters, that begin and end with a's and that have nothing but b's inside

ab*a = {aa aba abba abbba abbbba…}

- The language that contains all the strings of a's and b's in which all the a's come before all the b's

a*b* = {Λ a b aa ab bb aaa aab abb bbb aaa…}

Note:*(a*b* is not Equals to (ab)*)*

- The language of Σ = {a}, defining the odd language

a(aa)*or(aa)*a

- The language of Σ = {a}, defining the even language

(aa)*

- The language of the strings defined over Σ = {a,b}, which do not contain a double letter

b(ab)*a + a(ba)*b

Regular Expression cont…

- The language that contains all the strings of a's and b's of length = 3 exactly

  $(a+b)^3$or (a+b)(a+b)(a+b)
- The language of all word that have at least two a's

  (a+b)*a(a+b)*a(a+b)*
- The language of all strings with exactly two a's

  b*ab*ab*
- The language which denotes all the words with at least two a's

  (a+b)*a(a+b)*a(a+b)* = b*ab*a(a+b)*
- The language of all the words with exactly two b's or exactly two a's

  b*ab*ab* + a*ba*ba*

Regular Expression cont…

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

*need to understand carefully

- The language of all words that contain both an a and a b is defined by the expression

  (a + b)*a(a + b)*b(a + b)* + bb*aa*
- The set of all the strings of a's and b's that at some point contain a double letter

  (a+b)*(aa+bb)(a+b)*
- The language of string of even number of a's , followed by odd number of b's or even number of b's, followed by odd number of a's

  (aa) *b(bb) * + (bb) *a(aa) *
- The language of all words without a double a

  b*(abb*)a

Regular Expression cont…

- String of all words having exactly two a’s

  b*ab*ab*
- Language of all strings of words, starting with a and ending with b or starting with b and ending with a

  a(a+b)*b + b(a+b)*a
- The String of all words whose length(x) < 3, starting with ba

  ba
- The String of words, starting with double b and ending with either a or b

  bb(a+b)
- All the strings ending at aa or bb

  (a+b)*(aa+bb)or(a+b)*aa + (a+b)*bb

Regular Expression cont…

- Any string that contains at least two consecutive a’s
  (a+b)*aa(a+b)*

- All words that don’t ends at ba
  (a+b)*(aa+ab+bb)
- All words, starting with double letter
  (aa+bb)(a+b)*
- All words that contains at least one double letter
  (a+b)*(aa+bb)(a+b)*

- *All words that start and end with a different double letter
  aa(a+b)*bb(a+b)* + bb(a+b)*aa(a+b)*

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

Regular Expression cont…

- All words of length > 3
  (a+b)(a+b)(a+b)(a+b)(a+b)* or (a+b)(a+b)(a+b)(a+b)+
- All word of length >= 3
  (a+b)(a+b)(a+b)(a+b)* or (a+b)(a+b)(a+b)+
- All words that start and end with a double letter
  aa(a+b)*bb + bb(a+b)*aa or (aa+bb)(a+b)*(aa+bb)
- All words that contain exactly one a or one b
  b*ab*+a*ba*
- All words that contain exactly two a’s or exactly two b’s
  (a+b)*aa(a+b)* + (a+b)*bb(a+b)*
- The language of string in which any no. of a’s may occur before, between, and after the b’s
  a*ba*ba*

Regular Expression cont…

- All words that contains at least two a’s
  b*ab*ab*
- All words with at least two a’s
  b*ab*a(a+b)*
- All words with exactly two a’s
  b*ab*ab*
- Language that defines all even-length strings of alternating a’s and b’s
  (ab)*+(ba)*
-

  (a + b)(aa + bb + ab + ba)*
- The set of strings with an even number of a's followed by an odd number of b's
  (aa)*(bb)*b

Regular Expression cont…

- The set of strings over {a, b} that contains the substrings aa or bb

  (a+b)*aa(a+b)* + (a+b)*bb(a+b)*
- The set of string over { a, b } that do not contain the substrings aa and bb

  (ab)* - (ab)*aa(ab)* + (ab)*bb(ab)*
- The strings that start with aa, end with bb, and have alternating substrings ba in between

  $a(ab)^{+}$ boraa(ba)*bb
- The strings that contain at most one b and the rest a's

  a*(b+Λ)a*
- The even length strings of a's and b's

  (aa+bb+ab+ba)*
- The odd length strings of a's and b's

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

Regular Expression cont…

- The set of all strings of a’s and b’s that have at least two letters, that begin and end with a’s and that have only b’s in between

  ab*a = {aa,aba,abba,abbba,abbbba,...}
- All the words that begin with either an a or a c and then are followed by some number of b’s

  (a+c)b* = {a,c,ab,cb,abb,cbb,abbb,cbbb,abbbb,cbbbb,…}
- The only words that do not contain both an a and a b in them

  a*+b*
- All words of the form some positive number of a’s followed by exactly one b

  aa*b
- All words of strings that contain at most one b and the rest a's

  a*(b+Λ)a*

Regular Expression cont…

- The language, consists of strings that are either all a’s or b followed by a nonnegative number of a’s

  a*+ba*or (Λ+ b) a*

- The language that denote all words with at least two a’s

  b*ab*a(a+b)*

- The language of all words in which either the a come before b or the b come before the a

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

- The language of all strings of a's and b's that have even length

  (aa+ab+bb+ba)*or((a+b)(a+b))*

- *The set of all strings over {a, b, c} that do not contain the substring ac

  (c*(a+(bc*))*)

*Need to understand carefully

Regular Expression cont…

- Language of all words that have at least one *a* and at least one *b*

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

What about the word ***ba***

?

MUST BE=>

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

Regular Expression cont…

- All strings with prefix ab

  ab(a+b)*

- All strings with postfix ab

  (a+b)ab*

- All strings with prefix ba or prefix ab

ba(a+b)* + ab(a+b)*

- All strings with postfix ba or postfix ab

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

- All strings with prefix ab and postfix ba

ab(a+b)*ba

- All strings with prefix ba and postfix ab

ba(a+b)*ab

Prefix

Postfix

Regular Expression cont…

- All strings that have two consecutive a’s

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

- All strings except those with two consecutive a’s

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

- All strings with an even number of a’s

(b*ab*ab*)*

- Language of all even length strings of alternating a’s and b’s

  (ab)* + (ba)*

- The set of strings over {a,b} that end in 3 consecutive b's

  $(a + b)^*$ bbb

- The set of strings that have at least one b

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

Over View:

- The language L of strings of odd length, defined over Σ={a}, can be written as **L={a, aaa, aaaaa,…..}**
- The language L of strings that does not start with a, defined over Σ={a,b,c}, can be written as**L={b, c, ba, bb, bc, ca, cb, cc, …}**
- The language L of strings of length 2, defined over Σ={0,1,2}, can be written as **L={00, 01, 02,10, 11,12,20,21,22}**
- The language L of strings ending in 0, defined over Σ ={0,1}, can be written as **L={0,00,10,000,010,100,110,…}**
- The language EQUAL, of strings with number of a's equal to number of b's, defined over Σ={a,b} = **{Λ ,ab,aabb,abab,baba,abba,…}**
- The language EVEN-EVEN, of strings with even number of a's and even number of b's, defined over Σ={a,b}, can be written as

  {Λ, aa, bb, aaaa,aabb,abab, abba, baab, baba, bbaa, bbbb,…}

Over View:

- The language INTEGER, of strings defined over Σ={-,0,1,2,3,4,5,6,7,8,9}, can be written as **INTEGER = {…,-2,-1,0,1,2,…}**
- The language EVEN, of stings defined over Σ={-,0,1,2,3,4,5,6,7,8,9}, can be written as **EVEN = { …,-4,-2,0,2,4,…}**
- The language $\{a^n b^n\}$, of strings defined over Σ={a,b}, as $\{a^n b^n$: n=1,2,3,…}, can be written as **{ab, aabb, aaabbb,aaaabbbb,…}**
- The language $\{a^n b^n a^n\}$, of strings defined over Σ={a,b}, as $\{a^n b^n a^n$: n=1,2,3,…}, can be written as {aba, aabbaa, aaabbbaaa,aaaabbbbaaaa,…}
- The language factorial, of strings defined over Σ={1,2,3,4,5,6,7,8,9} *i.e.* {1,2,6,24,120,…}
- The language **FACTORIAL**, of strings defined over Σ={a}, as $\{a^{n!}$: n=1,2,3,…}, can be written as {a,aa,aaaaaa,…}

  It is to be noted that the language FACTORIAL can be defined over any single letter alphabet.

Over View:

- The language **DOUBLEFACTORIAL**, of strings defined over $\Sigma=\{a, b\}$, as $\{a^{n!}\ b^{n!} : n=1,2,3,\ldots\}$, can be written as {ab, aabb, aaaaaabbbbbb,…}
- The language **SQUARE**, of strings defined over $\Sigma=\{a\}$, as $\{a^{n^2} : n=1,2,3,\ldots\}$, can be written as {a, aaaa, aaaaaaaaa,…}
- The language **DOUBLESQUARE**, of strings defined over $\Sigma=\{a,b\}$, as $\{a^{n^2}\ b^{n^2} : n=1,2,3,\ldots\}$, can be written as {ab, aaaabbbb, aaaaaaaaabbbbbbbbb,…}
- The language **PRIME**, of strings defined over $\Sigma=\{a\}$, as

  $\{a^p : p \text{ is prime}\}$, can be written as {aa,aaa,aaaaa,aaaaaaa,aaaaaaaaaaa…}

- // Total Examples Defined over {a,b} = 100

Languages and Regular Expressions

Regular Expressioncont…

Regular Expression cont…

Note: $(a^* b^*)^* = (a + b)^*$

- Some important expressions equal to :

$$1^* (1 + \Lambda) = 1^*$$

$$1^* 1^* = 1^*$$

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

$$(0^* 1^*)^* = (0 + 1)^*$$

$$(0 + 1)^* 01(0 + 1)^* + 1^* 0^* = (0 + 1)^*$$

Regular Expression cont…

Write RE for the following languages for $\Sigma = \{0,1\}$

- The set of strings over {0,1} that end in 3 consecutive 1's.
  $(0 + 1)^* 111$
- The set of strings over {0,1} that have at least one 1
  $0^* 1 (0 + 1)^*$
- The language that consists of all strings where the length of any run of consecutive 0's is a multiple of 3

  $(1 + 000)^*$
- The language of all strings that end in 1101

  $(1 + 0)^*1101$
- Language that defines all even-length strings of alternating 0s and1s, where $\Sigma = \{0,1\}$

  $(01)^* + (10)^*$

Regular Expression cont…

- The language of those binary strings, that is, strings on the alphabet {0, 1}, that contain the substring 1011

  (0+1)*1011(0+1)*
- The language of all binary strings where every run of consecutive 1's has even length

  (0 + 11)*
- The language of all binary strings that do not contain the substring 1011

  ((0*11*00)*0*11*0(10)*0)*(0*+11*+11*0(10)*+11*0(10)*1)
- The set of all strings over {0,1} that do not have the substring 111

  (0*+(((0*(1+(11)))((00*)(1+(11)))*)0*))

Regular Expression cont…

- Language of all even length strings of alternating 0s and 1s
  (01)* + (10)*
- The language of all strings of 0's and 1's that have odd length
  (0 + 1)(00 + 01 + 10 + 11)*
- Set of all strings with any number of "0"s followed by any number of 1s
  0*1*
- All strings that have two consecutive 0s
  (0+1)*00(0+1)*
- All strings except those with two consecutive 0s
  (1*01)*1* + (1*01)*1*0

Regular Expression cont…

- All strings with an even number of 0s

  (1*01*01*)*
- All the strings of length 2

  $(00 + 01 + 10 + 11)^*$
- Language of all strings that ends in 1 and doesn't contain the substring 00

  (1+01)*
- All strings except those with two consecutive 0's

  (1*01)*1* + (1*01)*1*0
- All strings with an even number of 0's

  (1*01*01*)*

Regular Expression cont…

- Language of all even length strings of alternating 0's and 1's

  $(01)^* + (10)^*$
- The set of strings over {0,1} that end in 3 consecutive b's

  $(0 + 1)^*\ 111$
- The set of strings that have at least one 1

  $0^*1(0+1)^*$
- **All strings without substring 001**

?

Think & Do Your Self…

- Question: *Write a regular expression for the set of strings that contains an even number of 1's over □={0,1}. Treat zero 1's as an even number.*
- Answer:

(0+10)*11(0+1)*

Important Examples & Applications

Regular Expressionscont…

Example No. # 1

- Let L be the language of all strings of 0s and 1s that have even length, (Since 0 is even, L contains Λ). Is L regular, and if so, what is a regular expression corresponding to it?
- We can answer this by realizing that if a string has even length, it can be thought of as consisting of a number, possibly zero, of string of length 2 concatenated.
- And, conversely, any such concatenation has even length.
- Since we can easily enumerate the strings of length 2, we may write the answer:

$(00 + 01 + 10 + 11)^*$

Example No. # 2

- Let L be the language of all string of 0's and 1s that have odd length. We can use the previous example: odd length means in particular length at least one, and so we may view L as the language of all strings consisting of single symbol followed by an even-length string. Since we have a regular expression for even-length strings, and we can easily find one for strings of length 1, a regular expression for L is

  $(0 + 1)\ (00 + 01 + 10 + 11)^*$

- one may ask why we couldn't have described the language in this example as the set of string consisting of an even-length string followed by a single symbol, which would have led to

  $(00 + 01 + 10 + 11)^*\ (0 + 1)$

Example No. # 3

- Let L be the language of all strings of 0s and 1s containing at least one 1.
- Here are three regular expressions corresponding to L:

  -$0^* 1 (0 + 1)^*$

  -$(0 + 1)^* 1 (0 + 1)^*$

  -$(0 + 1)^* 10^*$
- The first expresses the fact that a string in L can be decomposed as follows: an arbitrary number of 0's (possibly none), the first 1, and then any arbitrary string.
- The second, which is some sense is the most general, or the closest to our definition of L, expresses the fact that a string in L has a 1, both preceded and followed by an arbitrary string.
- The third is similar to the first, but emphasized the last 1 in string in L.

Example No. # 4

- L = { w is a binary string which does *not* contain two consecutive 0s or two consecutive 1s anywhere)
  - e.g., w = 01010101 is in L, while w = 10010 is not in L
- Goal: Build a regular expression for L
- Four cases for w:
  - Case A: w starts with 0 and |w| is even
  - Case B: w starts with 1 and |w| is even
  - Case C: w starts with 0 and |w| is odd
  - Case D: w starts with 1 and |w| is odd
- Regular expression for the four cases:

  ==□

Case A: (01)*	Case B: (10)*
Case C: 0(10)*	Case D: 1(01)*

- Since L is the union of all 4 cases:
  - R.E for L = (01)* + (10)* + 0(10)* + 1(01)*
- If we introduce Λ then the regular expression can be simplified to:
  - R.E for L = **(Λ +1)(01)*(Λ +0)**

+

(Λ +0)(10)*(Λ +1)

Example No. # 5

- L = {x belongs to {0, 1} where x ends with 1 and does not contain the sub-string 00}
- This mean that every string in L corresponds to the regular expression R = $(1 + 01)^*$
- This extra constraint simply means that Λ can't be included, and that L corresponds to the regular expression.

  $(1 + 01)^+ = (1 + 01)^* (1 + 01)$

Assignment No. # 1

Very Important for Regular Expressions

Theory of Automata

Q.No.1. Write regular expression for each of the following language over the alphabet set {1, 0}

i) The language of all strings not containing the substring 000

ii) The language of all strings that do not contain the substring 110

iii) The language of all strings containing both 101 and 010 as substrings

iv) The language of all strings in which both the number of 0's and the number of 1's are even

v) The language of all strings in which both the number of 0's and the number of 1's are odd

vi) All words that contain exactly two 1's or three 1's, not more

vii) All strings that have exactly one double letter in them

viii) All words in which 1 is tripled or 0 is tripled, but not both. This means each word contains 111 or 000 but not both

ix) All strings in which the total number of 1's is divisible by 3 no matter how they are distributed, such as 1101100101

x) All words in which any 0's that occur are found in clumps of odd number at a time, such as 1101100010

Assignment No. # 1 (Solution)

Important R.E.

- **Write a R.E for set of strings over {a,b} that do not contain the substring *aa***

Regular Expression cont…

Note:**Regular expressions describe regular languages**

Example:$(a + bc)^*$

= {, a , bc , aa , abc , bca , ….}

// Total Examples Defined over {0,1} = 40

Regular Expression cont…
Example: (Solved the Un-Solved)

1. $01^* = \{0, 01, 011, 0111, \ldots..\}$
2. $(01^*)(01) = \{001, 0101, 01101, 011101, \ldots..\}$
3. $(0+1)^*$
4. $(0+1)^*01(0+1)^*$
5. $((0+1)(0+1)+(0+1)(0+1)(0+1))^*$
6. $((0+1)(0+1))^*+((0+1)(0+1)(0+1))^*$
7. $(1+01+001)^*(\Lambda+0+00)$

How to make and check Regular Expression Easily?

Regular Expressioncont…

Regular Expression cont…

Example $(a+b)\cdot a^*$

Regular expression:

$$
\begin{aligned}
L((a+b)\cdot a^*) &= L((a+b))\,L(a^*) \\
&= L(a+b)\,L(a^*) \\
&= (L(a)\cup L(b))\,(L(a))^* \\
&= (\{a\}\cup\{b\})\,(\{a\})^* \\
&= \{a,b\}\,\{\lambda,a,aa,aaa,\ldots\} \\
&= \{a,aa,aaa,\ldots,b,ba,baa,\ldots\}
\end{aligned}
$$

Exercise:

- Practice the Examples of Regular Expressions where Σ={ i , j , k }