Regular Expressions

Text Book &Resources:

Introduction to Computer Theory–2nd Edition - (I. O. Cohen)

Regular Expressions

Rules to Build Regular Expression

Definition

The set of **regular expressions** is defined by the following rules:

- Rule 1: Every letter of the alphabet Σ can be made into a regular expression by writing it in **boldface, ε** itself is a regular expression.
- Rule 2: If r_1 and r_2 are regular expressions, then so are:
  (i) (r_1)
  (ii) r_1r_2
  (iii) $r_1 + r_2$
  (iv) r_1*
- Rule 3: Nothing else is a regular expression.
- Note: If r_1 = aa + b then when we write r_1* , we really mean $(r_1)*$,

  that is $r_1* = (r_1)* = (aa + b)*$

Example9: The language start and ends on same letter defined over ∑ = {a, 1, 0}

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

Example10: The language whose second and second last letter are same defined over ∑= {a, 1, 0}

RE=$aa+00+11+(a+1+0)(a+1+0)(a+1+0)+$

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

Example11: The language contain substring ab or ba defined over, ∑= {a, b}

RE=$(a+b)^*ab(a+b)^*+(a+b)^*ba(a+b)^*$

Or

RE=$(a+b)^*(ab+ba)(a+b)^*$

Example12: The Language contain exact two a's defined over, ∑={a,b}

RE=b*ab*ab*

Example13: The Language contain at most two a's defined over, ∑= {a, b}

RE=b*+b*ab*+b*ab*ab*

Example14: The language cannot end on 'ab' defined over, ∑= {a, b}

RE=$(a+b)^*(ba+aa+bb)+\Lambda+a+b$

Example15: The language cannot contain even length of word defined over, ∑= {0, 1, 2}

RE=$((0+1+2)(0+1+2))^*(0+1+2)$

Example16: The language contain word of length multiple of 3 defined over, ∑= {a, b}

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

Example17: The language cannot contain substring 'aa' defined over, ∑= {a, b}

RE=$\Lambda+a+b+(ab+b)^*(\Lambda+a)$

Example18: The Language contain even length of word but not multiple of 3 defined over, ∑= {a, b}

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

$((a+b)(a+b)+(a+b)(a+b)(a+b)(a+b))$

Example19: The Language contain length of multiple of 3 but not multiple of 4 defined over, ∑= {a, b}

RE= $((a+b)(a+b)(a+b)+(a+b)(a+b)(a+b)(a+b)(a+b)(a+b)+$

$+b((a+b)(a+b)(a+b)(a+b)(a+b)(a+b)(a+b)(a+$

$a+ba+ba+ba+ba+ba+ba+ba+ba$

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

Example20: Language of all words having at least two a's

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

Example21: Language of all words having exact two a's

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

Example22: Language of all words having same starting and ending

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

Example23: Language of all words which start with 'ab' and end with 'bb'

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

Example24: Language of all words having even length $((a+b)(a+b))^*$

Example25: Language of all words which contain substring 'ab'

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

Example26: Language of all words having even number of a's and b's

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

Example27: Language which contain odd number of a's and b's

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

Green parts mention above will produce Λ or even numbers of $a's$ and $b's$ and yellow portion will produce ab or ba. As a whole (even-even) (ab + ba) (even - even) will produce odd number of $a's$ and odd number of $b's$.

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

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

Example28: Language which has length 4 and start and end with different letters

$$a((a+b)(a+b))b+b((a+b)(a+b))a$$

Example29: Language which has exactly one 'aa' or 'bb' and have length greater then 3

$$a(aa+bb)b+b(aa+bb)a+b(aa)b+a(bb)a$$

Example30: Language of words which may or may not contain a's or b's and all b's must be placed after a's

$$a_* b_*$$

Example31: Language which start with 'aa' and end with 'bbb' and length will be greater then 8

$$aa((a+b)^+(a+b)(a+b)(a+b))bbb$$

Example32: Language which start with 'a' and must contain 'bb'

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

Example33: Language which has exactly 1 'b' and must have length greater then 3

$$baaa^+ + abaa^+ + aaba^+ + aaa^+ba^* + aa^+ba^+$$

Example34: Language of all strings which has at least 1 'a' and 1 'bb' $(a + b)^*a(a+b)^*bb(a+b)^* + (a+b)^*bb(a+b)^*a(a+b)^*$

Example35: Language of all strings which start and end with same letters and length greater then 4

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

Example36: Language of all strings which has exact 1 'aaa' and must start with 'b'

$$b^+(aaa)b^*$$

Example37: Language which accept length >=2 and not end with 'aa' or 'bb'

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

Example38: Language which contain ‘aa’ or ‘bb’ and always end with ‘a’

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

Example39: Language which has odd length but start with ‘b’

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

Example40: Language which only accept ‘bba’or ‘bb’

$$bba+bb$$

Assignment

Produce regular expressions for the following?

Q#1

Password

$C = \{A\text{-}Z\}$

$S = \{a\text{-}z\}$

$D = \{0\text{-}9\}$

$X = \{\$, \_, \#\}$

Conditions:

- Password must start and end with small or capital letters.
- Password may or may not contain special characters.
- Password must have digits.
- Password must have length greater than 3.
- Password must not start with digit or special characters.

Q#2

Language of all those words which has two a’s and two b’s.