MEDTECH
CS420

Author : M. Taha BENNANI **Academic Year** : 2020-2021

Reminder: Arden's method

Following algorithm is used to build the regular expression from given DFA.

```
1. Let q1 be the inial state
2. There are q2, q3, ...qn number of states. The final state may
   be some qj where j<=n.
3. Let alpha(ji) represents the transition from qj to qi
4. Calculate qi susch that:
   qi = qj alpha(ji)
   if qi is a start state :
     qi = qj alpha(ji) + epsilon
5. Similarly compute the final state which gives the regular
   expression r
```

Exercice 1: Regular expressions

Find a regular expression for language over the alphabet 0,1 in which all the strings:

1. contain at least two 0's
2. have odd number of b's

Exercice 2: Regular expressions to Automatas

1. Draw the NFA for r1 = b, then draw the NFA for r2 = a*, and draw the NFA for r3 = ba*. Construct NFA for the regular expression b+ba*
2. Construct NFA for the regular expression (0+1)*
3. Construct NFA for the regular expression (01+10)+
4. Morse code is a way to encode letters using dots(.) and dashes (-). Here is the encoding of subset of letters that we consider in this problem.

Token name	Character	Token name	Character	Token name	Character
A	.-	L	.-..	T	-
E	.	R	.-.	P	.–.
I	..	V	...-	F	..-.

   (a) What is the alphabet used in this language?

(b) Create the automata that recognizes the above subset of Morse code tokens.

Exercice 3: NFA to DFA

1. Convert NFA the following NFA to its equivalent DFA

	0	1
-Q0+	Q3	Q1, Q2
Q1	Q4	Null
Q2	Null	Q3
Q3	Q3	Q4
Q4	Null	Null

2. Construct NFA for the regular expression (epsilon + (0+1) (00)*) Then generate its equivalent DFA.

Exercice 4: Automata to Regular expressions

1. Compute the regular expression recognized by the following automata:

	a	b
-Q1+	Q2	Q3
Q2	Q4	Q1
Q3	Q1	Q4
Q4	Q4	Q4

2. Construct the regular expression from the following DFA:

	a	b
-Q1	Q1	Q2
Q2+	Q3	Q3
Q3	Null	Null

3. Find out the regular expression from given DFA

	0	1
-Q1+	Q1	Q2
Q2	Q3	Q2
Q3	Q1	Q2

Exercice 5: Maximal Munch

1. Let L1 the language over the alphabet {a,b} defined by the following regular expressions:
   T1 = aa
   T2 = b?a+b?
   T3 = b?a*b?
   T4 = .| \n

(a) Generate the automata which recognizes L1

(b) Analyze the input "bbbabaa" and compute the differents lexemes

2. Let L2 the language over the alphabet {a,b,c} defined by the following regular expressions:
T1 = aa*b*c*
T2 = c*b*

(a) Generate the automata which recognizes L2

(b) Analyze the input "babcaababccbcabb" and compute the differents lexemes

(c) Consider the following tokenization: R1 = c, R2 = ac+b*, R3 = cc, R4 = ab*, R5 = a, R6 = ac*b+. Which tokens can never be generated and why?