

Assignment 2 – Lexical Analyser

SEG2106 – Software Construction

Question 1 – Regular Expressions (30 points – 6 points each)

Find regular expressions that define the following languages:

1. XML opening and closing tags where the opening tags may include attributes. The form of such tags is given by the following examples: `<name-x attribute-1="some string" attribute-2="xyz">` or `<funny laugh123="go" s-t-o-p=";>"/>` or `</name>`. The alphabet is composed of the sets alpha, digit and `{-,<,>,/,"",=,;,}`.

Regular expression to find the opening tag:

`(<\s*[\w\-_]+\s*([\w\-_]+="[^"]*"|'\s*'|\s+\/\s+)?\s*>)`

Regular expression to find the closing tag:

`(<\/\s*[\w\-_]+\s*>)`

Putting the two together:

`(<\s*[\w\-_]+\s*([\w\-_]+="[^"]*"|'\s*'|\s+\/\s+)?\s*>)|(<\/\s*[\w\-_]+\s*>)`

2. All strings over the alphabet `{a, b}` that do not contain the substring `aaa`.
`a{0,2}(b+a{0,2}b)*`
3. All strings over the alphabet `{a, b}` for which the number of "a" is a multiple of 3 (including zero).
`(b*(ab*ab*a)*)*`

4. All binary numbers greater than 10111.

We assume that there might be a leading zero.

$(0^*11(0|1)(0|1)(0|1))|$

$(0^*1(0|1)^*(0|1)(0|1)(0|1)(0|1)(0|1)(0|1))$

5. All strings of the kind EPX where E is an integer number, P is a lowercase letter and X is an integer greater than 3 and less than 13. Examples: 143a6, 555b12, etc.

To match the integer:

$(0|((-?) [1-9][0-9]^*))$

To match the lower case letter:

$[a-z]$

To match an integer greater than 3 and less than 13:

$((1[0-2])|[3-9])$

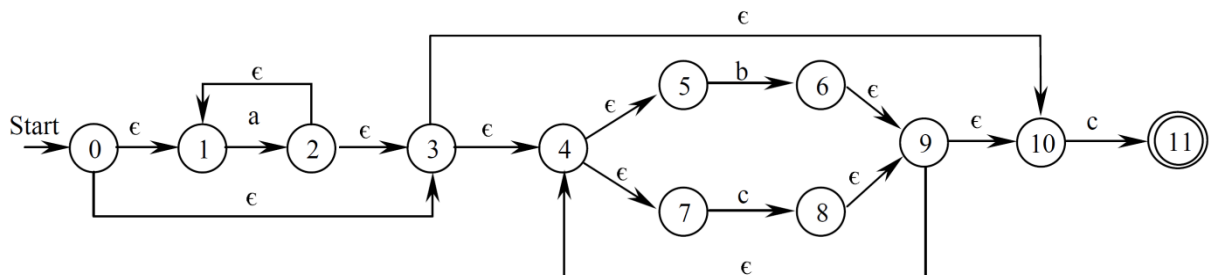
Putting it all together:

$(0|((-?) [1-9][0-9]^*)) [a-z] ((1[0-2])|[3-9])$

Question 2 – Non-Deterministic Finite Automata (30 points – 10 points each)

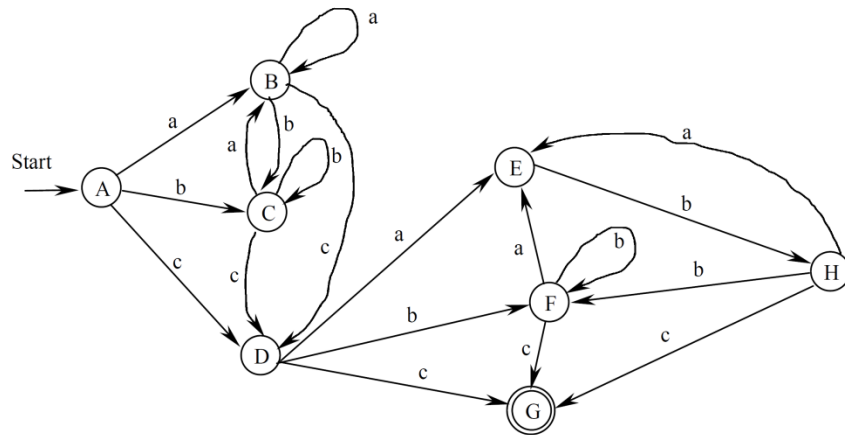
Convert the following regular expressions to Non-deterministic Finite Automata (NFA):

- a) $a^* (b | c)^* c$



- b) $((b|a)^*|(c|a)^*(cb)^*)$

$\epsilon\text{-closure}(\text{move}(A,b)) = \epsilon\text{-closure}(\{5\}) = \{1,2,4,5,6,7\} = C$
 $\epsilon\text{-closure}(\text{move}(A,c)) = \epsilon\text{-closure}(\{8\}) = \{8,9,12,15\} = D$
 $\epsilon\text{-closure}(\text{move}(B,a)) = \epsilon\text{-closure}(\{3\}) = \{1,2,3,4,6,7\} = B$
 $\epsilon\text{-closure}(\text{move}(B,b)) = \epsilon\text{-closure}(\{5\}) = \{1,2,4,5,6,7\} = C$
 $\epsilon\text{-closure}(\text{move}(B,c)) = \epsilon\text{-closure}(\{8\}) = \{8,9,12,15\} = D$
 $\epsilon\text{-closure}(\text{move}(C,a)) = \epsilon\text{-closure}(\{3\}) = \{1,2,3,4,6,7\} = B$
 $\epsilon\text{-closure}(\text{move}(C,b)) = \epsilon\text{-closure}(\{5\}) = \{1,2,4,5,6,7\} = C$
 $\epsilon\text{-closure}(\text{move}(C,c)) = \epsilon\text{-closure}(\{8\}) = \{8,9,12,15\} = D$
 $\epsilon\text{-closure}(\text{move}(D,a)) = \epsilon\text{-closure}(\{10\}) = \{10\} = E$
 $\epsilon\text{-closure}(\text{move}(D,b)) = \epsilon\text{-closure}(\{13\}) = \{8,9,12,13,14,15\} = F$
 $\epsilon\text{-closure}(\text{move}(D,c)) = \epsilon\text{-closure}(\{16\}) = \{16\} = G\text{-accepting}$
 $\epsilon\text{-closure}(\text{move}(E,a)) = \epsilon\text{-closure}(\{\}) = \{\}$
 $\epsilon\text{-closure}(\text{move}(E,b)) = \epsilon\text{-closure}(\{11\}) = \{8,9,11,12,14,15\} = H$
 $\epsilon\text{-closure}(\text{move}(E,c)) = \epsilon\text{-closure}(\{\}) = \{\}$
 $\epsilon\text{-closure}(\text{move}(F,a)) = \epsilon\text{-closure}(\{10\}) = \{10\} = E$
 $\epsilon\text{-closure}(\text{move}(F,b)) = \epsilon\text{-closure}(\{13\}) = \{8,9,12,13,14,15\} = F$
 $\epsilon\text{-closure}(\text{move}(F,c)) = \epsilon\text{-closure}(\{16\}) = \{16\} = G\text{-accepting}$
 $\epsilon\text{-closure}(\text{move}(G,a)) = \epsilon\text{-closure}(\{\}) = \{\}$
 $\epsilon\text{-closure}(\text{move}(G,b)) = \epsilon\text{-closure}(\{\}) = \{\}$
 $\epsilon\text{-closure}(\text{move}(G,c)) = \epsilon\text{-closure}(\{\}) = \{\}$
 $\epsilon\text{-closure}(\text{move}(H,a)) = \epsilon\text{-closure}(\{10\}) = \{10\} = E$
 $\epsilon\text{-closure}(\text{move}(H,b)) = \epsilon\text{-closure}(\{13\}) = \{8,9,12,13,14,15\} = F$
 $\epsilon\text{-closure}(\text{move}(H,c)) = \epsilon\text{-closure}(\{16\}) = \{16\} = G\text{-accepting}$



Transition Table:

| State/Input | a | b | c |
|-------------|---|---|---|
| A | B | C | D |
| B | B | C | D |
| C | B | C | D |
| D | E | F | G |
| E | - | H | - |
| F | E | F | G |
| G | - | - | - |
| H | E | F | G |

b)

$\epsilon\text{-closure}(1)=\{1,6,13\}=A$

$\epsilon\text{-closure}(\text{move}(A,c))=\epsilon\text{-closure}(\{2,7,17\})=\{2,7,17\}=B$

$\epsilon\text{-closure}(\text{move}(B,o))=\epsilon\text{-closure}(\{3,8,18\})=\{3,8,18\}=C$

$\epsilon\text{-closure}(\text{move}(C,o))=\epsilon\text{-closure}(\{4,9\})=\{4,9\}=D$

$\epsilon\text{-closure}(\text{move}(C,m))=\epsilon\text{-closure}(\{14\})=\{14\}=E$

$\epsilon\text{-closure}(\text{move}(C,d))=\epsilon\text{-closure}(\{19\})=\{19\}=F$

$\epsilon\text{-closure}(\text{move}(D,l))=\epsilon\text{-closure}(\{5,10\})=\{5,10\}=G\text{-accepting}$

$\epsilon\text{-closure}(\text{move}(E,p))=\epsilon\text{-closure}(\{15\})=\{15\}=H$

$\epsilon\text{-closure}(\text{move}(F,e))=\epsilon\text{-closure}(\{20\})=\{20\}=I\text{-accepting}$

$\epsilon\text{-closure}(\text{move}(G,e))=\epsilon\text{-closure}(\{11\})=\{11\}=J$

$\epsilon\text{-closure}(\text{move}(H,i))=\epsilon\text{-closure}(\{9\})=\{9\}=K$

$\epsilon\text{-closure}(\text{move}(H,u))=\epsilon\text{-closure}(\{16\})=\{16\}=L$

$\epsilon\text{-closure}(\text{move}(J,r))=\epsilon\text{-closure}(\{12\})=\{12\}=M\text{-accepting}$

$\epsilon\text{-closure}(\text{move}(K,l))=\epsilon\text{-closure}(\{10\})=\{10\}=N$

$\epsilon\text{-closure}(\text{move}(L,t))=\epsilon\text{-closure}(\{10\})=\{10\}=N$

$\epsilon\text{-closure}(\text{move}(N,e))=\epsilon\text{-closure}(\{11\})=\{11\}=J$

