Compiler Design

Lecturer: Samaneh Hosseinmardi

Fall Semester 1401-1402

Writing Assignment 1 Deadline 1401/08/24

1 Regular Expressions

Give regular expressions that generate each of the following languages. In all cases, the alphabet is $\Sigma = \{a,b\}$.

a. The language $\{w \in \Sigma^* \mid |w| \text{ is even}\}$. (5pts)

b. The language $\{w \mid w$ contains at least two a's, or exactly two b's $\}$. (5pts)

c. The language $\{w \in \Sigma^* \mid$ w contains a double letter $\}$. (A string contains a double letter if it contains aa or bb as a substring.) (5pts)

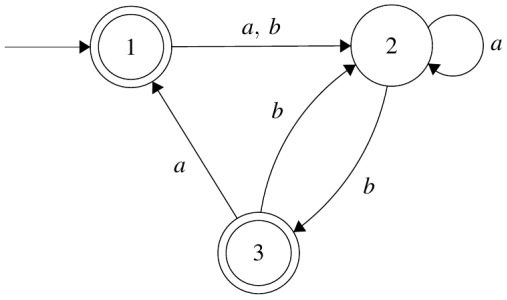
2 Minimum State DFA

Draw a DFA with the minimum possible number of states for recognizing multi-dimensional arrays. The name of arrays can be any string matching $[A-Za-z][A-Za-z0-9]^*$ and the dimensions of the array can only be digits ($[0-9]$). Note that you don't need to consider any garbage state in your design. (15pts)
Example: A3bT[5][6]

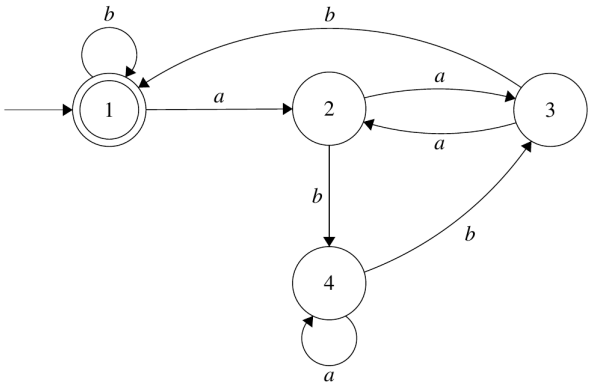
3 DFA to Regular Expression

Find a regular expression for each of the following DFAs.

a. $\Sigma = \{a,b\}$ (10pts)



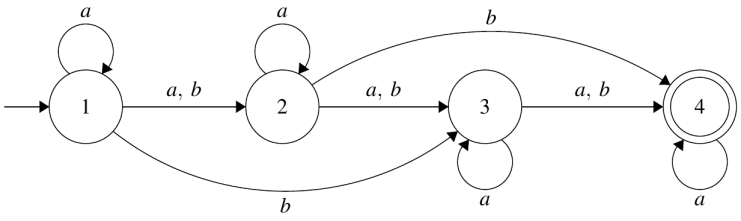
b. $\Sigma = \{a, b\}$ (10pts)



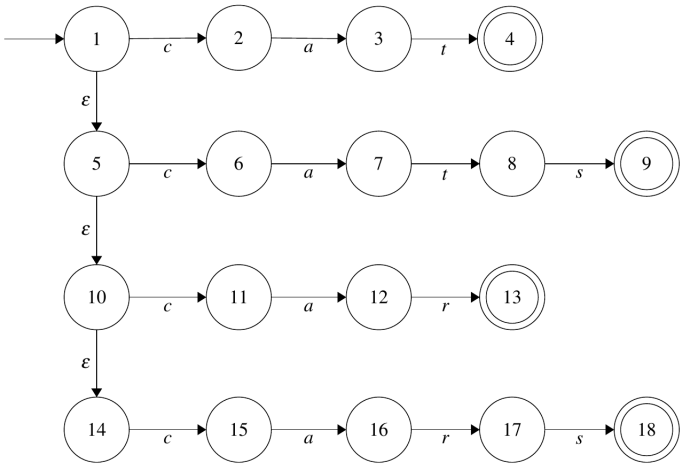
4 NFA to DFA

Convert the following NFAs into DFAs.

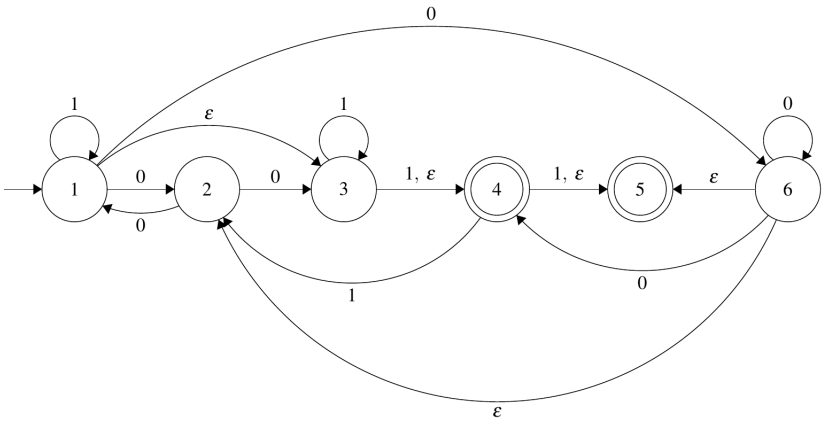
a. $\Sigma = \{a, b\}$ (10pts)



b. $\Sigma = \{a, c, r, s, t\}$ (10pts)



c. $\Sigma = \{0, 1\}$ (Extra credit: 10pts)



5 DFA

Consider the following set of token types: (30pts)

Token Type	Description
NUMBER	Any string matching: $[0-9]^+$
ID	Any string matching: $[A-Za-z][A-Za-z0-9]^*$
KEYWORD	if, else, void, int, while, break, switch, default, case, return, endif
SYMBOL	; : , [ ] ( ) {} + − ∗ = < == /
COMMENT	Any string between a /* and a */ OR any string after a // and before a \n or EOF[1]
WHITESPACE	blank (ASCII 32), \n (ASCII 10), \r (ASCII 13), \t (ASCII 9), \v (ASCII 11), \f (ASCII 12)

a. Draw appropriate DFAs (i.e., similar to the DFAs in pages 56-58 in Lecture note 3) for recognizing these tokens. Note that ID and KEYWORD are recognized by the same DFA, which is almost identical to the one on Page 57 of Lecture note 3!

b. In each of these DFAs, specify exactly what characters should be considered as compatible with the 'other' label. Note label 'other' in different DFAs are not necessarily referring to the same set of symbols.

c. Then combine these DFAs into a single DFA. The resultant DFA (provided that the DFA is correct and complete) can then be used as a flowchart for implementing a scanner in Programming Assignment 1.

Required Document

Please upload a zip or a pdf file in Quera(`https://quera.org/course/11991/`).

Deadline

Tuesday 23:59. 1401/08/24.

Contact Information

Ask your questions in Quera(`https://quera.org/course/11991/`)

Good Luck

[1] Please note that NUM, ID, and KEYWORD tokens can be the very last token of input programs and should be correctly recognized when we reach to EOF (End Of File).