COMP3173 Assignment 1 Lexical Analysis

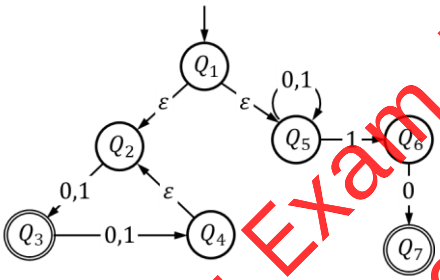
You may lose marks if your transition graphs are not clear.

1. Let $\Sigma = \{a, b\}$. Design a DFA for each following language (6 marks for each)
   a. $L_1 = \{ab^n a \mid n \geq 2\}$
   b. $L_2 = \{ba^n \mid n \geq 1 \text{ and } n \neq 3\}$
   c. $L_3 = \{w \mid |w| \bmod 3 \neq 0\}$, where $|w|$ is the length of the string w.

2. Design a regular expression for each language in Q1. (6 marks for each)

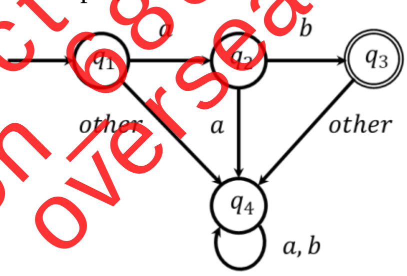
3. Let $\Sigma = \{0,1\}$. Convert the regular expression $(00)^*|(1(0|1)^*)$ to an NFA. (10 marks)

4. Given the following NFA



   a. Convert the NFA to the equivalent DFA. (20 marks)
   b. Minimize the DFA in Part a). (20 marks)

5. A ***trap state*** in a DFA is a state, which does not have any outgoing transitions (only looping back to itself). For example,



q_4 is a trap state in the DFA on the alphabet $\Sigma = \{a, b\}$. Once the DFA enters q_4, there is no way to go to the final state. Thus, in a real lexer, we can consider the trap state as a special "final state". Once a DFA enters a trap state, it reports a lexical error.

Furthermore, to simplify the DFA, we are allowed to use "$\boldsymbol{other}$" transition to take care of the undefined input symbol. For example, $q_1 \to q_2$ takes the symbol a. So, $other$ of q_1 is the undefined symbol b, because $\Sigma \setminus \{a\} = \{b\}$. Thus, $q_1 \to q_4$ takes the symbol b. Similarly, $q_3 \to q_4$ takes the symbol a or b.

Consider a new alphabet $\Sigma = \{a, \cdots, z\}$,

   a. Design a DFA with a trap state which recognizes the keywords "if" or "$else$". (The DFA does not accept any other string.) (10 marks)
   b. Represent the DFA using a transition table. (4 marks)

Note that errors may occur when the DFA stops on a non-final state or enters the trap state. You don't need to write down the regular expression and convert it to NFA and to DFA. The DFA can be constructed directly.