Midterm Exam

COSE215: Theory of Computation
2025 Spring

Instructor: Jihyeok Park

April 23, 2025. 13:30-14:45

- **If you are not good at English, please write your answers in Korean.**
  (영어가 익숙하지 않은 경우, 답안을 한글로 작성해 주세요.)
- **Write answers in good handwriting.**
  If we cannot recognize your answers, you will not get any points.
  (글씨를 알아보기 힘들면 점수를 드릴 수 없습니다. 답안을 읽기 좋게 작성해주세요.)
- **Write your answers in the boxes provided.**
  (답안을 제공된 박스 안에 작성해 주세요.)

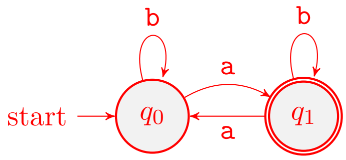
Student ID	
Student Name	

Question:	1	2	3	4	5	6	7	8	Total
Points:	15	25	15	5	5	10	10	15	100
Score:									

1. [15 points] Design **deterministic finite automata (DFA)** using **transition diagrams** that accept the following languages.

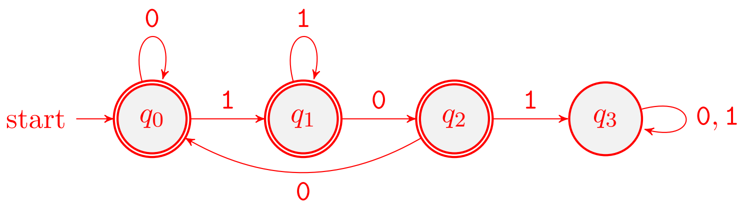
   (a) [5 points] $L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_{\mathtt{a}}(w) \equiv 1 \pmod 2\}$.

   Note that $N_{\mathtt{a}}(w)$ is the number of a's in w. (e.g., $N_{\mathtt{a}}(\mathtt{baab}) = 2$ and $N_{\mathtt{a}}(\mathtt{abaa}) = 3$.)

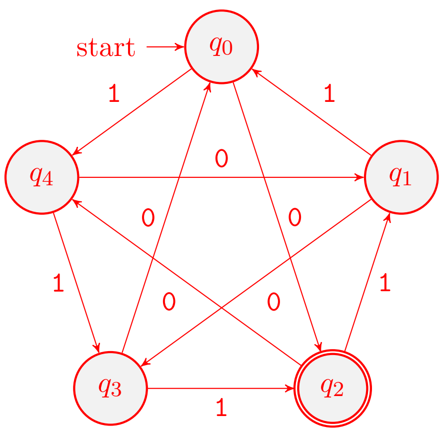


   (b) [5 points] $L = \{w \in \{0, 1\}^* \mid$ w does **not** contain 101 as a substring$\}$.

   Note that substrings are continuous sequences of symbols in a word (e.g., 101 is a substring of 010100.)



   (c) [5 points] $L \subseteq \{0, 1\}^*$ such that $h(L) = \{\mathtt{a}^n \mid n \equiv 2 \pmod 5\}$ where $h : \{0, 1\} \to \{\mathtt{a}\}^*$ is a homomorphism defined as $h(0) = \mathtt{aa}$ and $h(1) = \mathtt{aaaa}$.



2. [25 points] Write **regular expressions (REs)** that represent the following languages.

(a) [5 points] $L = \{w \in \{\texttt{a}, \texttt{b}\}^* \mid w \text{ has at most two } \texttt{a}\text{'s}\}$.

$R =$ $\texttt{b}^*\texttt{a}^?\texttt{b}^*\texttt{a}^?\texttt{b}^*$

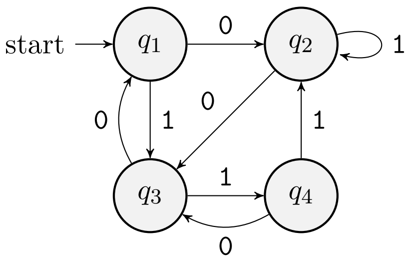
(b) [5 points] L is the reversal of the language defined by $\texttt{(ab|cd}^*\texttt{)}^*\texttt{ef}$ (i.e., $L = L(\texttt{(ab|cd}^*\texttt{)}^*\texttt{ef})^R$).

$R =$ $\texttt{fe(ba|d}^*\texttt{c)}^*$

(c) [5 points] $L = \{\texttt{a}^n\texttt{b}^m \mid n \times m \equiv 0 \pmod 3\}$.

$R =$ $\texttt{(aaa)}^*\texttt{b}^*\texttt{|a}^*\texttt{(bbb)}^*$

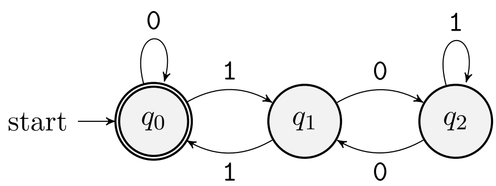
(d) [5 points] $L = L(R^{(3)}_{2,4})$ where $R^{(k)}_{i,j}$ be the regular expression that accepts the paths from q_i to q_j whose indices of the intermediate states are bounded by k in the following DFA in the left side:
(Hint: you can use the following regular expressions when $k = 2$ in the right side.)



$R^{(2)}_{1,1} = \epsilon$	$R^{(2)}_{1,2} = \texttt{0|01}^*\texttt{(}\epsilon\texttt{|1)}$	$R^{(2)}_{1,3} = \texttt{1|01}^*\texttt{0}$	$R^{(2)}_{1,4} = \varnothing$
$R^{(2)}_{2,1} = \varnothing$	$R^{(2)}_{2,2} = \texttt{1}^*$	$R^{(2)}_{2,3} = \texttt{1}^*\texttt{0}$	$R^{(2)}_{2,4} = \varnothing$
$R^{(2)}_{3,1} = \texttt{0}$	$R^{(2)}_{3,2} = \texttt{00|001}^*\texttt{(}\epsilon\texttt{|1)}$	$R^{(2)}_{3,3} = \epsilon\texttt{|01|001}^*\texttt{0}$	$R^{(2)}_{3,4} = \texttt{1}$
$R^{(2)}_{4,1} = \varnothing$	$R^{(2)}_{4,2} = \texttt{1|11}^*\texttt{(}\epsilon\texttt{|1)}$	$R^{(2)}_{4,3} = \texttt{0|11}^*\texttt{0}$	$R^{(2)}_{4,4} = \epsilon$

$R^3_{2,4} =$ $\texttt{1}^*\texttt{0(}\epsilon\texttt{|01|001}^*\texttt{0)}^*\texttt{1}$

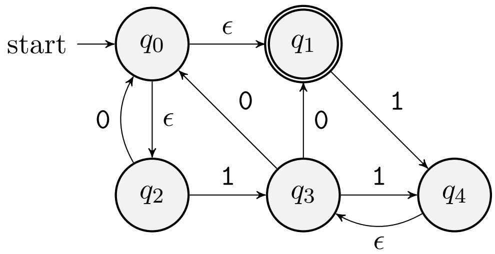
(e) [5 points] $L = h(L')$ where $h : \{\texttt{0}, \texttt{1}\} \to \{\texttt{a}, \texttt{b}\}^*$ is a homomorphism defined as $h(\texttt{0}) = \texttt{ab}$ and $h(\texttt{1}) = \texttt{a}$, and $L' \subseteq \{\texttt{0}, \texttt{1}\}^*$ is the language of the following DFA:



$R =$ $\texttt{(ab|a(aba}^*\texttt{ab)}^*\texttt{a)}^*$

3. [15 points] Consider the following **ϵ-nondeterministic finite automaton (ϵ-NFA)** N^ϵ:

$$N^\epsilon = (Q = \{q_0, q_1, q_2, q_3, q_4\}, \Sigma = \{0, 1\}, \delta, q_0, F = \{q_1\})$$



(a) [6 points] Construct a **DFA** D using a **transition table** such that $L(D) = L(N^\epsilon)$ via the **subset construction**.

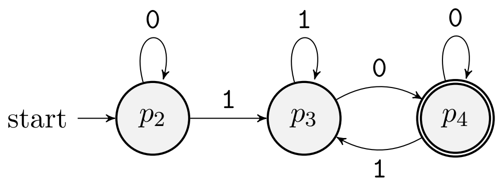
(Note that $\rightarrow$ indicates the **initial state**, and $*$ indicates a **final state**.)

$D =$

	$P = \mathcal{P}(Q)$	0	1
$* \rightarrow$	$p_0 = \{ q_0, q_1, q_2 \}$	p_0	p_1
	$p_1 = \{ q_3, q_4 \}$	p_0	p_1

(b) [6 points] Consider the following DFA $D' = (P' = \{p_2, p_3, p_4\}, \Sigma, \delta', p_2, \{p_4\})$:

$D' =$



Fill in the table in the left side using the **table-filling algorithm** for the states of D and D', and define the **equivalence classes** $(P \cup P')/_{\equiv}$ in the right side using the result of the algorithm.

p_1	X			
p_2	X	X		
p_3	X		X	
p_4		X	X	X
	p_0	p_1	p_2	p_3

$(P \cup P')/_{\equiv} = \{$

$\{ p_0, p_4 \}$,

$\{ p_1, p_3 \}$,

$\{ p_2 \}$,

$\}$

(c) [3 points] If D and D' are equivalent, **explain why**; otherwise, provide a word $w \in \{0, 1\}^*$ as a **counterexample** that produces different results from D and D'.

They are not equivalent because ϵ is accepted by D but not accepted by D'.

4. [5 points] Fill in the blanks in the **proofs** to show that $L = \{\mathtt{a}^j\mathtt{b}^k \mid 0 \leq 2j \leq 3k\}$ is **NOT regular**.

1. Assume that any positive integer n is given. (i.e., $n \geq 1$)
2. Pick a word $L \ni w =$ [$\mathtt{a}^{3n}\mathtt{b}^{2n}$].
3. $|w| =$ [$5n$] $\geq n$.
4. Assume that any split $w = xyz$ satisfying ① $|y| > 0$ and ② $|xy| \leq n$ is given.
5. Let $i =$ [2]. We need to show that ¬③ $xy^iz \notin L$:

Because of ②:

$$x = \mathtt{a}^p \qquad y = \mathtt{a}^q \qquad z = \mathtt{a}^{3n-p-q}\mathtt{b}^{2n}$$

for some $0 \leq p, q \leq n$ such that $p + q \leq n$.

Because of ①,

$$0 < q$$

Then, $xy^2z = \mathtt{a}^{3n+q}\mathtt{b}^{2n}$.

But, $2(3n + q) = 6n + 2q > 6n = 3(2n)$ ($\because 0 < q$).

So, $xy^2z \notin L$. □

5. [5 points] Consider a DFA $D = (Q, \Sigma, \delta, q_0, F)$ and a word $w \in L(D)$. Let m be the length of w. If $m \geq n$, we can split $w = xyz$ for some $q' \in Q$ satisfying the following conditions:

$$y \neq \epsilon \quad \land \quad \delta^*(q_0, x) = q' = \delta^*(q', y) \quad \land \quad \delta^*(q', z) \in F$$

For **what value of** n does the above statement always hold? And, **explain why**.

The value of n is the number of states in Q (i.e., $n = |Q|$).

Let $w = a_1 \cdots a_m$ and $p_i = \delta^*(q_0, a_1 \cdots a_i)$ for $0 \leq i \leq m$ where m is the length of w (i.e., $|w| = m \geq n$).

Then, because of the pigeonhole principle, there exists $i, j \leq n$ such that $p_i = p_j$ for some $i < j$.

Then, we can split w into three substrings $w = xyz$ such that

- $x = a_1 \cdots a_i$,
- $y = a_{i+1} \cdots a_j$,
- $z = a_{j+1} \cdots a_m$.

Because of the definition of δ^*, we have

- $\delta^*(q_0, x) = p_i = p_j = \delta^*(q_i, y)$,
- $\delta^*(q_i, z) \in F$.

So, we can conclude that $w = xyz$ satisfies the above statement. □

6. [10 points] Design **context-free grammars (CFGs)** that represent the following languages.

(a) [5 points] $L = \{\mathtt{0}^n\mathtt{1}^m \mid 0 \leq n \leq m\}$.

$$S \to \epsilon \mid \mathtt{0}S\mathtt{1} \mid S\mathtt{1}$$

(b) [5 points] $L = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid N_\mathtt{a}(w) = N_\mathtt{b}(w) + 1\}$.

Note that $N_\mathtt{a}(w)$ and $N_\mathtt{b}(w)$ are the number of a's and b's in w, respectively.

$$S \to ZS \mid \mathtt{a}Z$$
$$Z \to \epsilon \mid \mathtt{a}Z\mathtt{b} \mid \mathtt{b}Z\mathtt{a} \mid ZZ$$

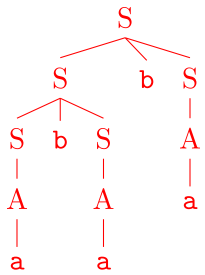
7. [10 points] Consider the following an **ambiguous CFG** G:

$$S \to A \mid S\mathtt{b}S \qquad A \to \mathtt{a} \mid \mathtt{a}A\mathtt{a}$$

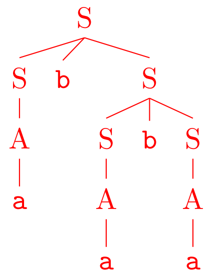
(a) [5 points] Pick a word $w \in L(G)$ that has two different parse trees and draw two parse trees of w:

$w =$ ababa

tree 1:



tree 2:



(b) [5 points] Convert G into an equivalent **unambiguous CFG** with **right-associativity** for b:

$$S \to A \mid A\mathtt{b}S$$
$$A \to \mathtt{a} \mid \mathtt{a}A\mathtt{a}$$

8. [15 points] A CFG G is called **right-linear** if its all production rules are of the form $A \to x$ or $A \to xB$ where $A, B \in V$ and $x \in \Sigma^*$.

(a) [9 points] Prove that the language of any **right-linear** CFG G is **regular** by designing a general algorithm that constructs an **ϵ-NFA** equivalent to the given G.

Let $V = \{A_1, A_2, \ldots, A_n\}$ be the set of nonterminals and A_1 be the start variable of G.

The core idea is to introduce a state q_i for each nonterminal A_i in G and a final state p.

Then, the initial state is q_1 and transitions are defined based on each production rule of G as follows:

- For $A_i \to a_1 a_2 \cdots a_k$, add transitions $q_i \xrightarrow{a_1} q'_1$, $q'_1 \xrightarrow{a_2} q'_2$, …, and $q'_{k-1} \xrightarrow{a_k} p$
- For $A_i \to a_1 a_2 \cdots a_k A_j$, add transitions $q_i \xrightarrow{a_1} q'_1$, $q'_1 \xrightarrow{a_2} q'_2$, …, and $q'_{k-1} \xrightarrow{a_k} q_j$

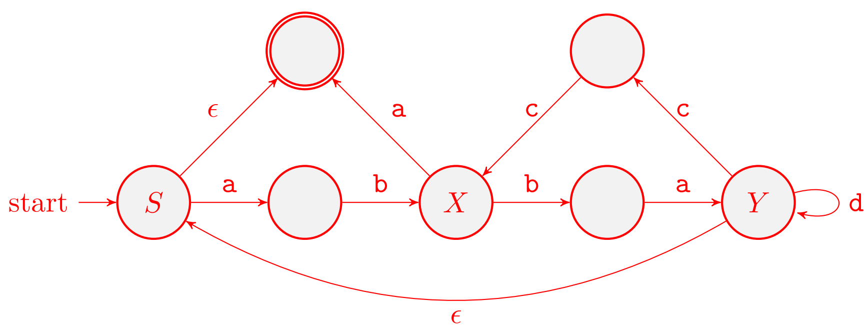
where q'_j is a new state for $1 \leq j < k$ for each production rule.

If $k = 0$, then add a transition $q_i \xrightarrow{\epsilon} p$ or $q_i \xrightarrow{\epsilon} q_j$ instead of the above transitions.

Then, the resulting ϵ-NFA mimics the behavior of G and accepts the same language as G. □

(b) [6 points] Construct an **ϵ-NFA** equivalent to the following **right-linear** CFG using the algorithm you designed in the previous question:

$$S \to \epsilon \mid \texttt{ab}X$$
$$X \to \texttt{a} \mid \texttt{ba}Y$$
$$Y \to S \mid \texttt{cc}X \mid \texttt{d}Y$$



This is the last page.
I hope that your tests went well!