Final Exam

COSE215: Theory of Computation
2025 Spring

Instructor: Jihyeok Park

June 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:	10	15	10	15	15	10	10	15	100
Score:									

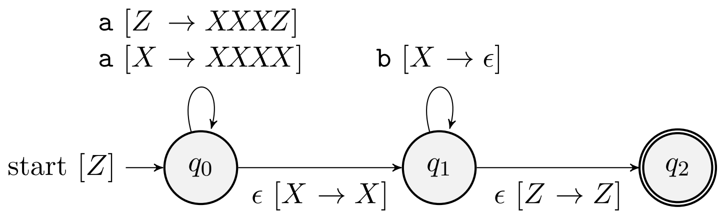
1. 10 points **True/False questions.** Answer O for True and X for False.

   1. The complement of a context-free language (CFL) is always a CFL. **X**
   2. All pushdown automata (PDAs) can be converted to equivalent context-free grammars (CFGs). **O**
   3. All deterministic CFLs can be defined by unambiguous CFGs. **O**
   4. The cardinality of the set of all Turing machines (TMs) is $\aleph_1$. **X**
   5. All CFLs can be recognized by empty stacks of a PDA having a single state. **O**
   6. If P = NP, then NP-hard problems are in P. **X**
   7. There is no polynomial-time reduction from a NP-hard problem to another problem. **X**
   8. There is no NP problem can be solved by a TM in polynomial time. **X**
   9. The universal language $L_u = \{(M, w) \mid M \text{ is a TM} \land w \in L(M)\}$ is undecidable. **O**
   10. Lambda calculus is a Turing-complete language. **O**

2. 15 points Explain the languages defined by the empty stacks (L_E) or final states (L_F) of PDAs P_i.

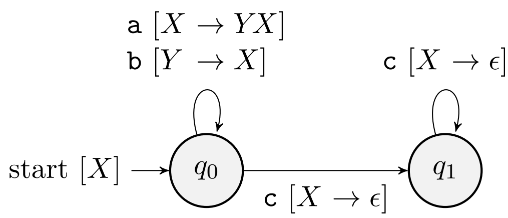
   (a) 5 points $L_F(P_1) = \{$ $\texttt{a}^n\texttt{b}^{3n} \mid n \geq 1$ $\}$

   $P_1 =$



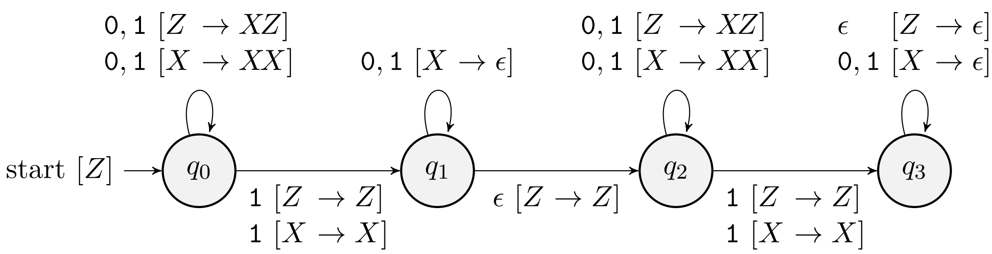
   (b) 5 points $L_E(P_2) = \{$ $(\texttt{ab})^n\texttt{c}^{n+1} \mid n \geq 0$ $\}$

   $P_2 =$



   (c) 5 points $L_E(P_3) = \{$ $w \in \{\texttt{0}, \texttt{1}\}^* \mid w = a_1 a_2 \cdots a_{2n} \land \exists i \in [1, n].\ a_i = a_{n+i} = \texttt{1}$ $\}$

   $P_3 =$

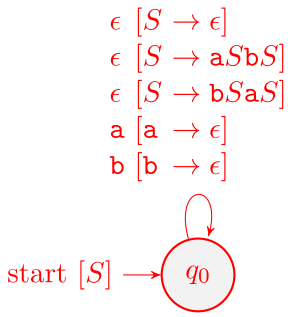


3. [10 points] We can freely **convert** a PDA to an equivalent **context-free grammar (CFG)** and vice versa.

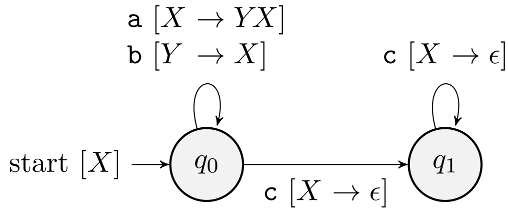
(a) [5 points] Consider the following CFG:

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

Construct a PDA accepting the language of the above CFG by its **empty stacks** with a **single state**.



(b) [5 points] Consider the following PDA:



Fill in the blanks in the production rules of the following CFG that represents the language accepted by **empty stacks** of the PDA.

S	$\to$	$A^X_{0,1}$
$A^X_{0,1}$	$\to$	$\mathtt{a}\ A^Y_{0,1}\ A^X_{1,1} \mid \mathtt{c}$
$A^X_{1,1}$	$\to$	$\mathtt{c}$
$A^Y_{0,1}$	$\to$	$\mathtt{b}\ A^X_{0,1}$

Note that each variable $A^X_{i,j}$ should generate all words that cause the PDA to move from the state q_i to the state q_j by popping the alphabet X.

$$A^X_{i,j} \Rightarrow^* w \qquad \text{if and only if} \qquad (q_i, w, X) \vdash^* (q_j, \epsilon, \epsilon)$$

Each right-hand side should consist of terminals and variables $A^X_{0,1}$, $A^X_{1,1}$, $A^Y_{0,1}$, $A^Y_{1,1}$ and might contain multiple productions. You can omit productions for useless (non-generating or unreachable) variables.

4. 15 points Consider the following languages:

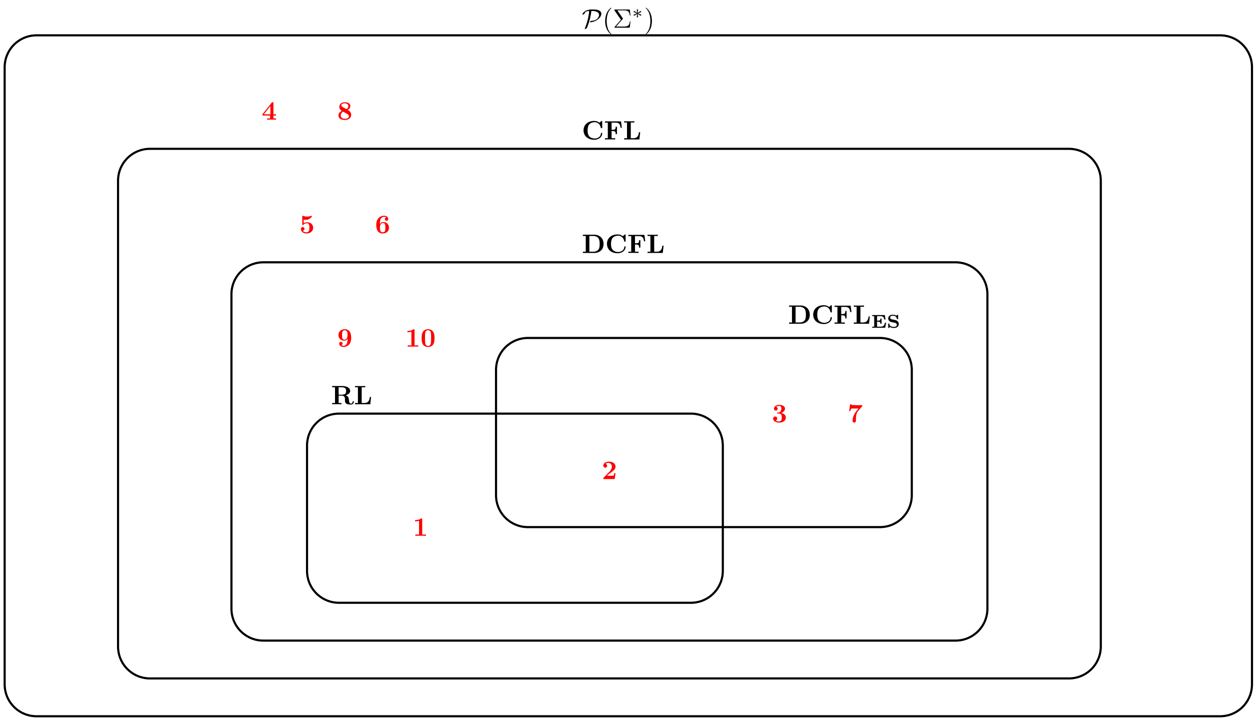
1. $L_1 = \{\mathtt{a}\mathtt{b}^n \mid n \geq 1\}$
2. $L_2 = \{\mathtt{a}^n\mathtt{b} \mid n \geq 1\}$
3. $L_3 = \{\mathtt{a}^n\mathtt{b}^n \mid n \geq 1\}$
4. $L_4 = \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \geq 1\}$
5. $L_5 = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid i = j \vee j = k\}$
6. $L_6 = \{ww^R \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$
7. $L_7 = \{w\mathtt{c}w^R \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$
8. $L_8 = \{w\mathtt{c}w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$
9. $L_9 = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid \mathbb{N}_\mathtt{a}(w) = \mathbb{N}_\mathtt{b}(w)\}$
10. $L_{10} = \{w \in \{\mathtt{a}, \mathtt{b}\}^* \mid \mathbb{N}_\mathtt{a}(w) \neq \mathbb{N}_\mathtt{b}(w)\}$

where $\mathbb{N}_\mathtt{a}(w)$ and $\mathbb{N}_\mathtt{b}(w)$ denote the number of a's and b's in w, respectively.

(a) 10 points Consider the following classes of languages.

- **CFL**: the class of **context-free languages**.
- **DCFL**: the class of **deterministic context-free languages**.
- **DCFL$_{\mathbf{ES}}$**: the class of **deterministic context-free languages by empty stacks**.
- **RL**: the class of **regular languages**.

The following Venn diagram shows the relationships between these language classes. Place the above languages in the following Venn diagram using their numbers (e.g., $1, 2, \ldots$).



(b) 5 points Some languages are **NOT DCFLs** but can still be defined by an **unambiguous** CFG. Find such a language in the **above list** and construct an **unambiguous CFG** for the language.

The language is $L_6 = \{ww^R \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$ and the CFG is:

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

5. 15 points Please construct new CFGs (G'_0, G'_1, and G'_2) from the given CFG (G_0, G_1, and G_2) by applying the following transformations.

(a) 5 points Construct a CFG G'_0 consisting of productions produced by replacing **nullable variables** with ϵ in all combinations and removing all ϵ-productions in production rules in G_0.

$$G_0 = \begin{cases} S \to AB \\ A \to \epsilon \mid \mathtt{0}A \\ B \to \mathtt{1} \mid AA \end{cases}$$

$$G'_0 = \begin{array}{l} S \to AB \mid B \mid A \\ A \to \mathtt{0}A \mid \mathtt{0} \\ B \to \mathtt{1} \mid AA \mid A \end{array}$$

(b) 5 points Construct a CFG G'_1 by removing all **unit productions** and adding all possible **non-unit productions** of Y to X for each **unit pair** (Y, X) in G_1.

$$G_1 = \begin{cases} S \to \mathtt{0} \mid A \mid B \\ A \to \mathtt{0}A \mid B \\ B \to \mathtt{1} \mid \mathtt{1}A \end{cases}$$

$$G'_1 = \begin{array}{l} S \to \mathtt{0} \mid \mathtt{0}A \mid \mathtt{1} \mid \mathtt{1}A \\ A \to \mathtt{0}A \mid \mathtt{1} \mid \mathtt{1}A \\ B \to \mathtt{1} \mid \mathtt{1}A \end{array}$$

(c) 5 points Construct a CFG G'_2 by removing all productions that contain **non-generating variables** or come from **unreachable variables** in G_2.

$$G_2 = \begin{cases} S \to \mathtt{0} \mid BD \mid \mathtt{1}E \\ A \to \mathtt{1} \mid AC \\ B \to \mathtt{0}D \mid \mathtt{1}BB \\ C \to \mathtt{0} \mid AB \\ D \to BD \\ E \to \mathtt{0}B \mid \mathtt{11} \end{cases}$$

$$G'_2 = \begin{array}{l} S \to \mathtt{0} \mid \mathtt{1}E \\ E \to \mathtt{11} \end{array}$$

6. 10 points Fill in the blanks in the **proof** showing that the language L is **not** a **context-free language (CFL)** using the **pumping lemma** for **CFLs**.

$$L = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid 0 \leq i \leq j \leq k\}$$

1. Assume that any positive integer n is given. (i.e., $n \geq 1$)
2. Pick a word $L \ni z =$ $\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n$.
3. $|z| =$ $3n$ $\geq n$.
4. Assume that any split $z = uvwxy$ satisfying ① $|vx| > 0$ and ② $|vwx| \leq n$ is given.
5. We need to show that ¬③ $uv^iwx^iy \notin L$ for some $i \geq 0$:

- vx cannot cover both a's and c's at the same time. ($\because$ ② $|vwx| \leq n$)
- If vx covers a's or b's but does not cover c's,
  - Let $i = 2$.
  - Then, uv^2wx^2y has at least one more a or b than c's. ($\because$ ① $|vx| > 0$).
  - Thus, $uv^2wx^2y \notin L$.
- Otherwise, vx covers b's or c's but does not cover a's.
  - Let $i = 0$.
  - Then, $uv^0wx^0y = uwy$ has at least one less b or c than a's. ($\because$ ① $|vx| > 0$).
  - Thus, $uv^0wx^0y \notin L$.

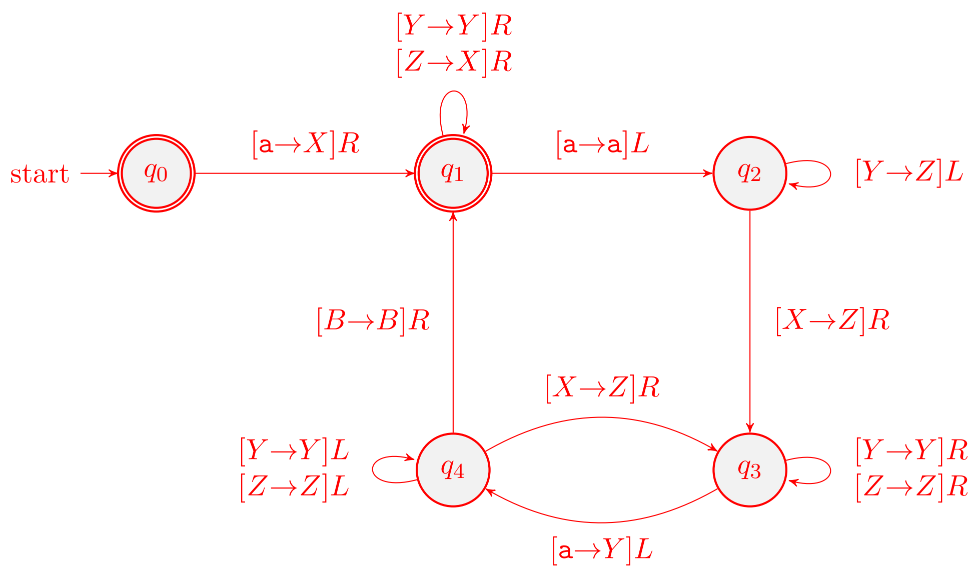
7. 10 points Draw a **Turing machine (TM)** M **accepting** the following language:

$$\{\mathtt{a}^n \mid n \text{ is a Fibonacci number}\}$$

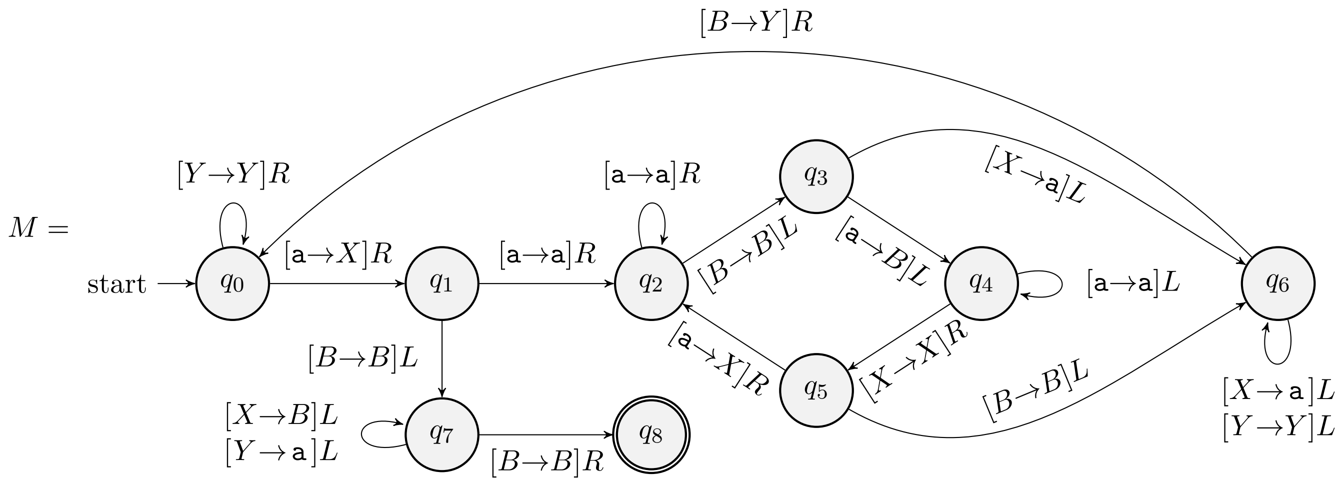
The k-th Fibonacci number $F(k)$ is defined as follows:

$$F(0) = 0, \quad F(1) = 1, \quad F(k) = F(k-1) + F(k-2) \text{ for } k \geq 2$$

$M =$



8. [15 points] Consider a **computable function** f defined by the following **TM** M.



(a) [2 points] The TM M takes an input $w \in \{\mathtt{a}\}^*$ and produces an output $f(w)$ only if $q_0\ w \vdash^* q_f\ f(w) \nvdash$ where q_0 is the initial state, q_f is one of the final states. Then, what is the **domain** (i.e., the set of all inputs for which the TM M produces an output) of the function f?

$\text{dom}(f) =$ $\{\mathtt{a}^n \mid n \geq 1\}$

(b) [8 points] What are the outputs of the function f for the following inputs?

$f(\mathtt{a}^2) =$ $\mathtt{a}^1$ $f(\mathtt{a}^5) =$ $\mathtt{a}^3$ $f(\mathtt{a}^{13}) =$ $\mathtt{a}^4$

Then, explain the general form of the output of the function f for any valid input $\mathtt{a}^n \in \text{dom}(f)$.

$f(\mathtt{a}^n) =$ $\mathtt{a}^{\lceil \log_2 n \rceil}$

(c) [5 points] Explain why the **time complexity** of the TM M is $O(n^2)$.

For each k-th cycle, the TM M requires the following moves:

1. The number of remaining $\mathtt{a}$'s is $n/2^k$.
2. $O((n/2^k)^2)$ moves to remove half of the remaining $\mathtt{a}$'s.
3. $O(k)$ moves to the leftmost cell to add one more Y.

After at most $\lceil \log_2 n \rceil$ cycles, the TM M replaces all Y's with $\mathtt{a}$'s in $O(k)$ moves and halts.

Therefore, it require the following number of moves:

$$\sum_{k=1}^{\lceil \log_2 n \rceil} O\left(\left(\frac{n}{2^k}\right)^2 + k\right) = O\left(\sum_{k=1}^{\lceil \log_2 n \rceil} \left(\frac{n^2}{4^k} + k\right)\right) = O(n^2) + O(\log^2 n) = O(n^2)$$

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