CS 501: Formal Language Theory	**Spring 2021**

Final Examination

Released: 5/12/2021, 11:30 am EST   Time Limit: 120 minutes   Due: 5/12/2021, 1:30 pm EST

Note: LaTeX template courtesy of UC Berkeley EECS dept.

Instructions. This final contains seven questions, for a total of 100 points (plus 5 bonus extra credit). You have a total of 120 minutes to complete it. You then have a grace time of 10 minutes to upload your final work in a single PDF document to Gradescope. *We will not accept any work submitted after that, so please plan accordingly.*

The final is an **individual effort**. You are required to write your entire attempt yourself, and are forbidden from consulting anyone else. Failure to abide by this will result in immediate failure from the course, among other consequences.

- You are **allowed** to consult the textbook, any of the lecture notes, and any material posted publicly for COMPSCI 501 this semester. If necessary, you are allowed to ask questions on the given Zoom link, or post on Piazza **privately** to the course staff for **clarifications** (bearing in mind that late posts have no guarantee of being answered). You are allowed to refer to your previous homeworks, the posted solution keys, and the discussion sheets and their solutions. If you happen to accidentally use any external online resources in your solutions, you are **required** to clearly acknowledge and cite this source.
- You are **not allowed** to search online for solutions, post on any public or private forum such as StackOverflow, or ask for help from a resource such as Coursera or Chegg. You are not allowed to collaborate or communicate in any form with anyone else other than course staff, in person or on chat (e.g. Slack, Discord, etc). You are also not allowed to post publicly on Piazza or on the Zoom public chat with a question that could give away part of an answer. You are not allowed to discuss the exam with anyone else until Friday, May 14, 11:59 pm.
- If you are **uncertain** about any of the rules, you should immediately seek clarification from one of the instructors who are proctoring.

Submissions. Please upload a PDF file before the deadline. Please start each new question on a new page. *Please ensure you tag your pages correctly, otherwise we cannot grade your work.*

---

1. **(10 × 2 points) Unjustified True/False Questions.** For each of the following questions, indicate simply whether it is TRUE or FALSE. *No justification necessary.*
   (a) The binary language $\mathsf{DOUBLED\text{-}SUBSTRING} = \{w : w = xyyz$, for some $x, z \in \{0,1\}^*$ and $y \in \{0,1\}^+\}$ is non-regular.
   (b) The language of any 10-state NFA also has some DFA with exactly 2021 states.
   (c) If L is any language, we define the language $\mathsf{expand}(L)$ to be $\{u : v \in L$ and u can be made by *inserting* any number of characters into v, in any order$\}$. Then if L is regular, so is $\mathsf{expand}(L)$.
   (d) If L is any language, we define the language $\mathsf{contract}(L)$ to be $\{u : v \in L$ and u can be made by *deleting* any number of characters from v, including leaving it alone$\}$. Then if L is context-free, so is $\mathsf{contract}(L)$.

(e) Let G be a CFG. Then there exists some number p such that any $w \in L(G)$ with $|w| > p$ can be written as $uvxyz$ such that v and y are both nonempty and that for any two integers $i, j \geq 0$, the string $uv^ixy^jz \in L(G)$.

(f) The language $\mathsf{START\text{-}STRING} = \{\langle G, w\rangle : G \text{ is a CFG that generates a string with prefix } w\}$ is decidable.

(g) By Rice's Theorem, the language $\mathsf{MACHINE\text{-}LANGUAGE} = \{\langle M\rangle : L(M) \text{ is Turing-recognizable}\}$ is undecidable.

(h) Let W be a finite set of words, each with two or more letters. A grid consists of black and white squares, with the letters on the white squares. Any interval of two or more consecutive white squares must form a word in the list (**you are allowed to repeat them**). Then,

$$\mathsf{XWORD\text{-}REPEATS} = \{\langle n, k\rangle : \text{ you can fill an } n \times n \text{ grid this way with } \leq k \text{ black squares}\}$$

is in PSPACE.

(i) The problem from the previous part is in PSPACE if you are **not allowed to repeat words**.

(j) A directed graph G is *pairwise accessible* if for every pair of distinct vertices u and v, there is some closed walk in G that starts from u, goes through v, and comes back to u. The language $\mathsf{PAIRWISE\text{-}ACCESSIBLE} = \{\langle G\rangle : G \text{ is a pairwise accessible directed graph}\}$ is PSPACE-complete.

2. **(5 $\times$ 6 points) Justified True/False Questions.** For each of the following questions, indicate whether it is TRUE or FALSE, *and provide a brief justification (i.e. either a proof or a counterexample).* **Assume throughout Q2 that P $\neq$ NP, and L $\neq$ NL.**

(a) Let $\mathsf{2021SAT} = \{\langle \phi\rangle : \phi \text{ is a Boolean formula with at least 2021 different satisfying assignments}\}$. Then, $\mathsf{2021SAT}$ is NP-complete.

(b) Consider the variant of GEOGRAPHY where a node can be visited multiple times (so draws are possible). Then, $\mathsf{EASY\text{-}GEOGRAPHY} = \{\langle G\rangle : \text{the first player wins in this variant}\}$ is in P.

(c) If NP $\neq$ PSPACE, then there *cannot* exist a problem X that is both PSPACE-complete and NP-complete.

(d) The problem $\mathsf{ACYCLIC\text{-}REACH} = \{\langle G, s, t\rangle : G \text{ is a directed acyclic graph with an } s\text{-}t \text{ path}\}$ is NL-complete.

(e) The problem $\mathsf{GRAMMAR\text{-}SUBSET} = \{\langle G, G'\rangle : L(G) \subseteq L(G')\}$ is in PSPACE.

3. **(5XC + 5 points) Product Placement.** Consider n matrices $A_1, \ldots A_n$, each of them $n \times n$, with the entries integers between -2^n and 2^n. Consider the problem of computing their n-way matrix product (recall that the product operation on $n \times n$ matrices is associative).

(a) (Bonus: 5 extra credit) Show that the entries of the product have only poly(n) bits.

(b) Assuming part (a) is true, place this problem in the smallest parallel complexity class you can.

4. **(10 points) Symbolic Gesture.** Consider the language

$$\mathsf{LETTER\text{-}REACHABILITY} = \{\langle M, a\rangle : \ M \text{ is a TM, } a \text{ is a tape symbol of } M\text{, and there exists some input word } w \text{ such that at some point } M \text{ writes } a \text{ on input } w\}$$

Show that $\mathsf{LETTER\text{-}REACHABILITY}$ is Turing-recognizable but not decidable.

5. **(10 points) Separate Checks.** For edges e, f in an undirected graph G, call them *separated* if there is no edge $g \in E(G)$ connecting an endpoint of e with an endpoint of f. Define the language

$$\mathsf{SEPARATED\text{-}EDGES} = \{\langle G, k\rangle : G \text{ is a graph with } k \text{ pairwise separated edges}\}.$$

Prove that $\mathsf{SEPARATED\text{-}EDGES}$ is NP-complete.

Hint: Create additional edges in the graph that do not share endpoints – you might double the size of the graph, but that's okay. Remember that $\mathsf{INDEPENDENT\text{-}SET}$ *is NP-complete as well.*

6. **(10 points) Accept the Emptiness.** Assume that E_{CFG} is P-complete. Use this fact to prove that A_{CFG} is P-complete as well.

7. **(6 + 9 points) JFK to BOS.** Let S be a *substitution system*, a fixed set of rules that can replace specific substrings of length 3 by others (i.e. substituting specific *trigrams*). Let

$$\begin{aligned}\mathsf{TRIGRAM\text{-}SUBSTITUTION} = \{\langle u, v, S\rangle :\ & u \text{ and } v \text{ are strings with } |u| = |v|, S \text{ is substitution system,} \\ & u \text{ can be transformed into } v \text{ by a finite sequence of } S\text{-moves}\}\end{aligned}$$

   (a) Prove that $\mathsf{TRIGRAM\text{-}SUBSTITUTION}$ is in PSPACE.

   (b) Prove that $\mathsf{TRIGRAM\text{-}SUBSTITUTION}$ is PSPACE-hard, by reducing from an arbitrary PSPACE problem.

   Example: if $S = \{\mathtt{OWL} \to \mathtt{HUM}, \mathtt{COO} \to \mathtt{COW}, \mathtt{CAT} \to \mathtt{DOG}, \mathtt{CHU} \to \mathtt{WAR}\}$, *then we have* $\langle \mathtt{COOL}, \mathtt{WARM}, S\rangle \in$ $\mathsf{TRIGRAM\text{-}SUBSTITUTION}$ *because of the sequence* $\mathtt{COOL} \to \mathtt{COWL} \to \mathtt{CHUM} \to \mathtt{WARM}$. *Note that of course in general, any strings in the relevant underlying alphabet can be used, English or not.*