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

Midterm Examination

Released: 3/26/2021, 10:30 am EST Time Limit: 90 minutes Due: 3/26/2021, 2:30 pm EST

Note: LaTeX template courtesy of UC Berkeley EECS dept.

Instructions. This midterm contains seven questions, for a total of 100 points. You have a total of 90 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 midterm 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 discuss the exam with anyone else until Monday, March 29, 11:59 pm.
- 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 problems publicly on Piazza until a week later.
- 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) Define
   $$L = \{w \in \{0,1\}^* : \text{some prefix of } w \text{ has exactly two more 1s than 0s}\}.$$
   Then L is a regular language that can be recognized with a DFA with 2021 or fewer states.

   (b) For a language L, define $L^{++} = \{w : w \text{ is a permutation of } v \text{ for some } v \in L\}$. That is, L^{++} is the language of permutations of all words in L. If L is a CFL, then L^{++} *must* also be a CFL.

   (c) Let $L = \{w \in \{a,b,c\}^* : w = a^n b^{2n} c^n, n \geq 0\}$. Then L is *not* context-free.

   (d) Let $L = \{\langle M, k\rangle : M \text{ is a TM that halts in at most } k \text{ steps on any input}\}$. Then L is decidable.

   (e) Let $L = \{\langle N\rangle : N \text{ is an NFA, and there is some input } w \text{ such that } N \text{ can end up in at least two distinct accept states on processing } w\}$. Then L is decidable.

(f) Let $L = \{\langle P \rangle : P$ is a PDA, and there is some input w whose computation makes the stack height at least 2021 at some point$\}$. Then L is decidable.

(g) Let $L = \{\langle D \rangle : D$ is a DFA, there are distinct strings u and v accepted by D, with u a prefix of $v\}$. Then L is decidable.

(h) Let $L = \{\langle M \rangle : M$ is a TM, there exists a DFA D with at most 2021 states such that $L(D) = L(M)\}$. Then L is decidable.

(i) The TD languages form a *single* equivalence class under the $\equiv_m$ relation.

(j) The Post Correspondence Problem is decidable over a one-character alphabet Σ.

2. **(5 × 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).*

(a) Let L be a regular language over $\{0, 1\}$, and let $L' = \{w' : |w'| = |w|$ for some word $w \in L\}$. Then L' *must* be regular.

(b) There is a language L that can be recognized by a DFA with 2020 states, but not by any DFA with 2021 states.

(c) The language of all even-length strings in $\{0, 1\}^*$ with at least one 1 in their second half is a CFL.

(d) A Turing machine with a tape that is infinite both to the left and to the right (initially filled with blanks except for the finite input) recognizes the same class of languages as ordinary TMs.

(e) A *special-substitution* operation replaces each letter $a \in \Sigma$ by a *string* $f(a)$ over the alphabet Γ. We extend the function f to strings in the obvious way. Then, if X is a CFL, then so necessarily is $\{w : f(w) \in X\}$.

3. **(5 points) Hanging TMs.** A *hanging TM* is one that hangs (i.e. immediately exits its computation) if it ever tries to move left from the leftmost cell of the tape. Prove that A_{HTM}, the accepting problem for hanging TMs, is undecidable.

4. **(10 points) Replicating Nonterminals** A nonterminal symbol A of a CFG G is called *replicating* if there exist strings $u, v \in (\Sigma \cup V)^*$ such that A eventually derives uAv. (Here V is the set of nonterminal symbols used in G.)

(a) Give an algorithm to test whether a specific non-terminal is replicating.

(b) Prove that if there are no replicating non-terminals, then $L(G)$ is regular.

5. **(10 points) Complementary Check-In.** For a language L, a *complement-enumerator* is a TM M that prints out all the strings in the *complement* of L in lexicographical order. Prove that a language L is decidable if and only if it has a complement-enumerator.

6. **(10 points) Reversal in Fortunes.** Let

$$\mathsf{ReverseCFL} = \{\langle L \rangle : L \text{ is a CFL and } L = L^R\}.$$

Prove that ReverseCFL is undecidable.

7. **(3 × 5 points) A Resolved Issue.** Fix a regular language L and let D be its minimal DFA. A string x is *unresolved* for L if there is both a string $xy \in L$ and a string $xz \notin L$; otherwise we say it is *resolved*.

(a) Prove that if two strings go to the same state in D, they are either both resolved or both unresolved.

(b) How can you determine from D which states correspond to unresolved strings?

(c) Prove that the number of states for resolved strings is either 0, 1, or 2.