Computer Science C63 | January 26, 2022
Scarborough Campus | University of Toronto

Homework Assignment #3
Due: February 9, 2022, by 11:59 pm

• ***You must submit your assignment through the Crowdmark system.*** *You will receive by email an invitation through which you can submit your work in the form of separate PDF docucments with your answers to each question of the assignment. To work with a partner, you and your partner must form a group on Crowdmark. Crowdmark does not enforce a limit on the size of groups.* ***The course policy that limits the size of each group to at most two remains in effect:*** *submissions by groups of more than two persons will not be graded.*
• *It is your responsibility to ensure that the PDF files you submit are legible. To this end, I encourage you to learn and use the LaTex typesetting system, which is designed to produce high-quality documents that contain mathematical notation. You are not required to produce the PDF files you submit using LaTex; you may produce it any way you wish, as long as the resulting document is legible.*
• *By virtue of submitting this assignment you (and your partner, if you have one) acknowledge that you are aware of the policy on homework collaboration for this course.*[a]
• *For any question, you may use facts previously proved in this course, its prerequisites, or in the assigned sections of the textbook.*
• *Unless we explicitly state otherwise, you should justify your answers. Your paper will be graded based on the correctness a of your answers, and the clarity, precision, and conciseness of your presentation.*

[a] "In each homework assignment you may collaborate with at most one other student who is currently taking CSCC63. If you collaborate with another student on an assignment, you and your partner must submit only one copy of your solution, with both of your names. The solution will be graded in the usual way and both partners will receive the same mark. Collaboration involving more than two students is not allowed. **For help with your homework you may consult only the instructor, TAs, your homework partner (if you have one), your textbook, and your class notes. You may not consult any other source.**"

In your answers for this assignment you may assume (by Church's thesis) that any algorithm described in pseudocode that involves no magical thinking (i.e., statements like "a miracle happens and ...") can be implemented by a Turing machine. Thus, to argue that a certain language L is decidable it suffices to present, in pseudocode, an algorithm that takes a string x as input and returns 1 if $x \in L$ and 0 if $x \notin L$. Similarly, to argue that L is recognizable it suffices to present, in pseudocode, an algorithm that takes a string x as input and returns 1 if $x \in L$, and returns 0 or does not terminate if $x \notin L$. Your pseudocode can use any common programming language features, and made-up higher-level constructs that are obviously translatable to pseudocode — e.g., "run Turing machine M on input x", "for every pair $(i, j) \in \mathbb{N} \times \mathbb{N}$ in dovetailing order do...".

Question 1. (20 marks) A set is closed under an operation if when the operation is applied to (the appropriate number of) elements of the set, the result is also an element of the set. For example, $\mathbb{N}$ is closed under addition, but not under subtraction. The set of decidable languages is closed under complementation, but the set of recognizable languages is not. For each statement below state whether it is true and justify your answer.

a. The set of decidable languages is closed under concatenation.

b. The set of recognizable languages is closed under concatenation.

c. The set of decidable languages is closed under the Kleene star operation. (Question to think about but not to include in your answer: The set of recognizable languages is closed under the Kleene star operation.)

Question 2. (10 marks) Twin primes are a pair of prime numbers that differ by two; e.g., 3 and 5, 11 and 13, 17 and 19, 137 and 139. The famous ***twin prime conjecture*** asserts that there are infinitely many such pairs. Nobody has succeeded yet in proving this, however. It is possible that such pairs cease to exist after some point; if so, this point must be very large indeed: according to Wikipedia, the largest twin primes known as of 2018 are the numbers $2{,}996{,}863{,}034{,}895 \cdot 2^{1{,}290{,}000} \pm 1$. But who knows, maybe these are the last two, and we run out of twin primes after that.

Consider the following two sets (recall that if $n \in \mathbb{N}$, $\langle n \rangle$ denotes an encoding of n):

$$A = \{\langle n \rangle : \ n \text{ and } n+2 \text{ are both primes}\}$$
$$B = \{\langle m \rangle : \ \text{for all } n \geq m,\ n \text{ and } n+2 \text{ are } \underline{\text{not}} \text{ both primes}\}$$

For each of A and B state whether the set is decidable, and prove your answer.

Question 3. (15 marks) An ***enumerator*** for a language $L \subseteq \Sigma^*$ is a multitape TM E_L that, if started with all tapes being blank, lists on Tape 1 every element of L preceded and followed by a special symbol #, where $\# \notin \Sigma$, and lists nothing else.

For an informal description, see pp. 180-181 in Sipser. More precisely, the computation of E_L started in its initial state with all tapes blank is a sequence of configurations $C_1 \vdash C_2 \vdash \cdots \vdash C_i \vdash \cdots$, such that
(a) if $i < j$ then the string contained in Tape 1 in C_i is a prefix of that in C_j; and
(b) for every $x \in \Sigma^*$, $\#x\#$ is a substring of the string contained in Tape 1 in some C_i if and only if $x \in L$.

The following fact provides an alternative characterization of recognizable languages:

Fact. (Sipser, Theorem 3.1) *Language L is recognizable if and only if there is an enumerator E_L for L.*

a. Prove that a language L is decidable if and only if there is an enumerator E_L for L that lists the strings in L in lexicographic order.[1]

b. Prove that every infinite recognizable language has an infinite decidable subset. (Hint: You may find useful the characterization of decidable languages that you proved in part (a).)

Question 4. (10 marks) Recall that $\leq_T$ is the Turing-reduces relation and $\leq_m$ is the mapping-reduces relation between languages. Let

$$U = \{\langle M, x \rangle : \ \text{Turing machine } M \text{ accepts input } x\}$$
$$H = \{\langle M, x \rangle : \ \text{Turing machine } M \text{ halts on input } x\}$$

a. Prove that $U \not\leq_m \overline{H}$ (i.e., U does <u>not</u> mapping-reduce to the complement of H).

b. Prove that $U \leq_T \overline{H}$.

c. Prove or disprove: For all languages L and L', if $L \leq_T L'$ and L' is recognizable, then L is recognizable.

Note: Parts (a) and (b) of this question imply that $\leq_m$ is <u>strictly</u> stronger than $\leq_T$: Although for all languages L, L', $L \leq_m L'$ clearly implies $L \leq_T L'$, the converse is not always true. Here is a question to think about, but not to include in your answer: A different way to prove that $L \leq_T L'$ does not necessarily imply $L \leq_m L'$ is to note that (a) for all languages L, $L \leq_T \overline{L}$ (why?) but (b) there is some language L such that $L \not\leq_m \overline{L}$ (can you give an example of such a language and prove that it has this property?).

[1] String $a_1 a_2 \ldots a_m$ appears before $b_1 b_2 \ldots b_n$ in lexicographic order if and only if $m < n$ or $m = n$ and for some i, $1 \leq i \leq m$, $a_j = b_j$ for all j, $1 \leq j < i$, and a_i is before b_i in the order of the symbols in alphabet Σ.