University of Waterloo
CS 462 — Formal Languages and Parsing
Winter 2020
Problem Set 11

Distributed Friday, March 27 2020. Due Friday April 3 2020 at 5 PM. Submit via LEARN.

This is the last problem set! Congratulations. You've made it through this course in challenging times. Celebrate and stay safe.

1. [10 marks] There is a slightly different version of Kolmogorov complexity $K(x)$, which is defined as the length of the shortest program-input pair (P, i) such that running P on i gives x, *but* demanding that the encoding of the pair (P, i) be *prefix-free*: no encoding is a prefix of any other. The point of this is that if we concatenate two such encodings $(P, i)(Q, j)$ we can always tell where the first ends and the second begins.

   For this version of Kolmogorov complexity prove that $K(xy) \leq K(x) + K(y) + O(1)$, where the constant in the big-O does not depend on x or y.

2. [10 marks] Let r be an arbitrary regular expression and G be an arbitrary grammar. Let $L(r)$ (resp., $L(G)$) be the corresponding languages. Which of the following two problems is solvable and which is unsolvable? Justify.

   (a) Decide if $L(r) \subseteq L(G)$;

   (b) Decide if $L(G) \subseteq L(r)$.

3. [10 marks] The point of this exercise is to show that concatenation can dramatically decrease Kolmogorov complexity.

   Let $C(x)$ denote Kolmogorov complexity. Give an example of a family of binary strings x_n and y_n such that $C(x_n y_n)/\min(C(x_n), C(y_n))$ tends to 0 as $n \to \infty$. Justify.