Homework 8:

Due: November 11, 2025 at 2:30p.m.

This homework must be typed in LaTeX and submitted via Gradescope.

Please ensure that your solutions are complete, concise, and communicated clearly. Use full sentences and plan your presentation before your write. Except where indicated, consider every problem as asking for a proof.

Problem 1. In document similarity analysis, the *shingling* algorithm is used to represent a document as a set of contiguous sequences (called *shingles*) to capture its structure and content. A k-character shingle uses contiguous substrings of length k characters, while a k-word shingle uses contiguous sequences of k words.
You are given the following two short documents:

- **Document A:** "The quick brown fox jumps over the lazy dog."
- **Document B:** "A quick brown dog jumps over a lazy fox."

1. Generate the set of **3-character shingles** for both documents (ignore punctuation and use lowercase).
   - List the first 10 distinct shingles for each document.
2. Generate the set of **3-word shingles** for both documents.
   - Show all 3-word shingles for each document.
3. Compute the **Jaccard similarity** between the two documents using:

   label=(i) 3-character shingles

   lbbel=(ii) 3-word shingles
4. Discuss the **differences** between using k-character and k-word shingles for text similarity.
   - Which method is more sensitive to small changes in wording or punctuation?
   - Which method better preserves semantic meaning?
   - Which would you choose for near-duplicate web page detection, and why?

Problem 2.

1. Let $F^k = \{f : \{0,1\}^k \to \{0,1\}\}$ be the set of all Boolean functions on k Boolean inputs. Give a very simple argument that the set F^k is countable.
2. Let $F^* = \cup_{i=1}^{\infty} F^k$. Show that the set F^* is countable.
3. Prove that all decision questions on finite graphs are decidable.
4. Let F^∞ be the set of Boolean functions with a countable number of Boolean inputs. Prove that F^∞ is uncountable.
5. (*) Give an example of a function in $F^\infty \setminus F^*$