Theory of Computation

Homework 2: due 6 May 2025

You are to write programs that implement

- LZ78 compression (encoding) and
- LZ78 decoding.

for normal English texts. You may assume that the alphabet Σ is $\{a, b, \ldots, z, A, B, \ldots, Z, 0, 1, \ldots, 9, ?, !\}$ plus the blank character, the comma, the period, the colon, the semicolon, and the newline character.

Use a trie for the dictionary of the LZ78 compression.

Compare the efficiency and compression ratio of your implementation with those by commonly used compression programs such as Winzip for various types of data. Show the results in your report. When you do encoding, try to obtain a best compression ratio.

- Run your program with your own input. An example running consists of
  - print input (e.g., `infile.txt`) of encoding
  - run the encoding of your program, and print encoding time and file sizes
  - run the decoding of your program, and print decoding time
  - print output (e.g., `outfile.txt`) of decoding
  - `diff infile.txt outfile.txt`
- Hand in your report, programs, and an example running (with at least two inputs).
- Write down the environment you run your program.
- Write comments appropriately in your program.