CSE 4094 Special Topics in Computer Engineering
Advanced Data Structures
Project 2
Mini Search Engine
Due: 24.01.2021 11:59PM

The purpose of this assignment is to use the suffix trees in a real-world application.

In this project, you will implement an application, in Java, C or Python; that takes a set of text files as input. The files shall be in a folder, of which the path is taken from the user.

The program will preprocess the given text files and answer the queries accordingly. The program shall not be case-sensitive, e.g., for a query including "data", it should also give output for the word "Data".

The queries can be:

- Which files include words starting with "xyz". Output shall also include the positions (i.e., the character number) of the given word in the corresponding files.
- What are the common word(s) in the files "f1", "f2", and "f3"?

In the implementation, you can use either of the following data structures: trie, compressed trie, suffix tree. You should discuss your choice of data structure and why you chose it in your report.

You will submit the source code of your program and your report using the Canvas service. The report shall include at least the following:

- A discussion about the selected data structure.
- Sample input and outputs for queries.

This is a group project. The groups can have 3 members at most.