CMP3005 Analysis of Algorithms
Term Project

Fall 2019

Due date: December 16, 2019 until 23:59

Introduction

In this project you will construct a system that fills the blanks in a given statement using the input text.

Submission

Submit all source code through **itslearning**. In addition to the source code, you should submit a min 3 pages report in a separate file which includes the methods (string matching method) you have applied, the programming language you have used, the libraries you have used, the average speed of your algorithm in terms of second, etc. The system will automatically be closed at the specified deadline: December 16, 2019 at 23:59 and the submissions after that time **will NOT be accepted**.

1 Implementation

1.1 Input files

Two files will be used by your algorithm: One which includes a text, and one containing statements which contain blanks at some location.

While implementing your project, you can use the files provided. The the_truman_show_script.txt file is the text through which you will be searching to fill in the blanks in the given statement. The statements.txt file contains example statements with some missing parts. The missing parts are represented with 3 underscore characters, i.e. ___ to which you must find answers from the text. Each statement is on a new line with a question mark at the end.

During the evaluation of your submission, the the_truman_show_script.txt file will be used as the text, however, the given statements are exemplary questions and it will not be the file you are graded on during your presentation.

1.2 Details

- A long text and a list of statements will be given as input to your program. For each statement, an answer must be found in the text and the statement should be printed to the screen as a complete sentence filled properly from the text.
- If the statement does not appear in the file, print "Statement NOT found" on the screen.
- You can use any text searching algorithm you would like, you can even use algorithms not discussed in class. The pattern matching algorithm must be written by yourself. You will be graded on the speed of your code, so you should try to choose an efficient algorithm.

Example

For example, say the text below will be searched for an answer:

> Pink Floyd were an English rock band formed in London in 1965. Gaining a following as a psychedelic band, they were distinguished for their extended compositions, sonic experimentation, philosophical lyrics and elaborate live shows, and became a leading band of the progressive rock genre. There are hit albums, and then there’s Dark Side of the Moon. Pink Floyd‘s eternally popular song cycle has sold more than 15 million copies in the U.S. since its release on March 1st, 1973, and more than 45 million units worldwide.

Given the input statement: they were distinguished for their extended compositions, sonic ____, philosophical lyrics and elaborate live shows

Your algorithm should output: experimentation

1.3 Important Instructions

When your program is executed, it will automatically fill in the blanks for each statement in the statements.txt file using the_truman_show_script.txt file as main text. It will NOT get any other inputs.

Your program should give an output to the console in the following format when executed:

1) {Original statement with missing part}
{Completed statement}
2) {Original statement with missing part}
{Completed statement}
.
n) { Original statement with missing part }
{ Completed statement }

1.4 Grading

You will be graded on the run time of your algorithm, so it is expected that your algorithm runs efficiently. Also, the accuracy of your program will be taken into account and will be measured as the percentage of statements your algorithm correctly completed from the text.

35% Running time, efficiency

20% Accuracy of the answers

20% Report

25% Algorithms used, code structure

1.5 Submission Instructions

- You must work in groups of 2 people. Individual or group more than 2 students will NOT be accepted.
- You will present your project at the last week of the semester and explain all details of the methods you used.
- Implementation can be done with C++, Java, or Python.
- Do NOT use built-in text search functions of the programming language you used.
- Submit your source code and report through itslearning as a .zip file.

- Only one person out of each group should submit the project. The file name should include all group members' student numbers in the format {STUDENT NUMBER} and {STUDENT NUMBER}.zip.

Cheating Policy

Cheating is strictly prohibited. Everything must be your own work, do not use each other's source code. If cheating is confirmed all students involved **will be penalized heavily.**

Important: itslearning has built-in plagiarism control that automatically detects submitted material that is plagiarised.