Programming Assignment

Submission deadline: 17th April 2022.

1. Write a program that takes an initial webpageurl as an input and a number. It then finds all out links of the webpage and traverses the links in breadth first traversal till the shortest distance between a webpage and the initial webpage is within the given number.
   It outputs website url, heading and parent url in an excel file. There should be no repetition. A webpage should be occurring only once in the output. [10]

2. Write a program that takes input a phrase, a webpage and a number, n; and then outputs all the webpages that match the phrase (exact phrase match), are within 'n' distance from the input webpage. The order in which output is displayed should be according to cosine similarity between the initial phrase and the webpage content. The output is should be stored in excel format by the program. [10]

Note:

1. Each assignment should be done using python.
2. There is no negative marking.
3. Copying is not allowed. There should be no plagiarism. Negative marks will be awarded if submitted assignments match.
4. If a program doesn't run, marks are zero.
5. Add a few proper comments in the program.