

Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23

5CS462 : PE5 - Data Mining Lab

Assignment No. 8

Use / extend the data analysis tool (menu driven GUI) developed in Assignment No. 1 to perform the following Web Mining task :

1. Implement the Crawler (DFS & BFS). The program should able to take seed input as any URLs/page and display all the links obtained by crawler.
2. Implement the PageRank algorithm to calculate the rank of each page in the file. The output should be the 10 pages with the highest rank, together with their rank values.
 - a. Tabulate the results containing adjacency matrix and rank of pages.
3. Implement the HITS algorithm to calculate the hub and the authority weight of each web page in the data set. The output should be the 10 most authoritative pages and 10 most hubby pages.
 - a. Tabulate the results containing adjacency matrix and rank of pages.

Note : Use the following data set from UCI machine learning repository :

- a. Web Graph datasets from “Stanford Large Network Dataset Collection”
<http://snap.stanford.edu/data/#web>

Deadline : 15/11/2022

Dr. B. F. Momin
Course Teacher