

CSC 395 : Information Retrieval
Fall 2019
Assignment 01
Building a Mini-Search Engine
Full Credit: 100

Assigned Date: 09/10/2019
Mid-Progress Presentation: 09/17/2019
(In class, 6 min presentation/testing)
Final Submission Due: 09/24/2019 (10:30 PM)

Problem Description:

Web search is by far the most common application involving information retrieval. We have discussed different components of search engine in class. Now we plan to implement a mini-search engine. We are free to choose any platform (programming language, tools, etc.). At the end, we will submit the followings:

- [90 points] A directory with fully documented source code
- [10 points] A README (instructions to execute your file, specifications of prerequisites)

Your submission will get full credit if it does at least the followings:

- a. Server Side should be able to crawl at least 100 unique pages, parse them, create index, build web-graph by analyzing the links, and rank them (you may want to implement page-rank)
- b. It is a good idea to design a topical search engine. Like, crawl pages whose title contains a particular topic: “sports”. That way, your web-graph will be well connected.
- c. Client Side should have a very simple interface to take input (text query) from the user, search in the index, and print the related web-links in some order.

We will talk more about it in class.