HW Due Tuesday Oct 2 11:59 pm

Homework 1 Distributed Systems. Concurrent Web Crawler

Simple Distributed Web Crawler. A Web crawler starts with a URL. As the crawler visits the URL, it identifies all the hyperlinks in the page and adds them to the list of URL to visit, called the crawl frontier. URLs from the frontier are recursively visited using BreadthFirstSearch. By now we will just print the links visited

Find all the links that can be reached from an initial -input variable- link and print them to the screen. The program must terminate, add depth limiting to the concurrent crawler. If user enters -depth=3, then only URLs reachable by at most three links will be fetched

You can find a sequential implementation of the crawler attached with this assignment

TODO: write a concurrent program of the above web crawler. Time your implementation and compare it to the sequential code provided

Upload to Polylearn :

Code (only your concurrent implementation part)

Table with the times (sequential execution vs concurrent) As a pdf document

During Office hours or Lab time you will Demo quickly the code running to me . Office hours are 11-12. and then lab is 12.10-1.30. I will be in the lab all that time and start looking at code in Alphabetical Last name order, you only get a full grade if you show me the code to me on that day