Women in STEM Project Proposal

Questions:

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

Kathleen Barta	kbarta2@illinois.edu	(Captain of the team)
Christine Zhou	xizhou4@illinois.edu	
Sharon Lin	xinyiyl2@illinois.edu	
Kris Kamaei	mkamaei2@illinois.edu	
Payel Chakraborty	payelc2@illinois.edu	

2. What system have you chosen? Which subtopic(s) under the system?

We have chosen the Intelligent Browsing theme for our project. Specifically, the subtopic we will implement is a browser extension called StudyFriend. The extension will scrape and index course related pages on sites such as Coursera and Campuswire, and then a user can search for a topic in the extension and be presented with links to related content on all available sites. This will streamline the process for students searching for material by eliminating the need to search on each site.

3. Briefly describe any datasets, algorithms or techniques you plan to use:

The dataset we will be using is the available pages on Coursera, Campuswire, and CMT. The algorithms used will include a web scraper algorithm and a similarity search algorithm. Possible algorithms include keras or NLTK for document analysis and classification. We will also be leveraging a browser extension development and bookmarking features if possible.

4. If you are adding a function, how will you demonstrate that it works as expected? If you are improving a function, how will you show your implementation actually works better?

The tool's function can be demonstrated by a user searching for a topic in the browser extension. Then the extension will return pages that are relevant to the topic. If time permits for further development, the user could bookmark useful pages.

5. How will your code communicate with or utilize the system? It is also fine to build your own systems, just please state your plan clearly

The extension will run on the Chrome browser. The user is expected to be authenticated prior to use in order to effectively scrape the pages from Coursera, etc, that require a login. Then the code will run on demand when the user activates the extension and enters a topic.

6. Which programming language do you plan to use?

For this project, we plan to use Javascript for the browser extension front end, using React, JSON, and atom.io as applicable. Then we will use python for the back end of the project, for the web scraping and indexing portions.

7. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

    a. Build a browser extension - 15 hours
    b. Build a modal that will display results - 15 hours
    c. Build a scraper - 15 hours
    d. Build an index and ranking - 20 hours
    e. Build a bookmark feature - 15 hours
    f. Test end to end and integration - 10 hours
    g. Documentation - 5 hours
    h. Create a presentation + tutorial - 5 hours