Progress Report

Project: Chrome Extension – Search Campuswire and Coursera

#	Task	Status	Any challenges/issues being faced.
1	Dataset: Write a crawler to get Coursera URLs of the course.	Complete	The App needs user to login, the program will open browser instance and wait for user to enter login credentials. Cannot programmatically login as Coursera website uses captcha.
2	Dataset: Write a crawler to get Campuswire URLs of a course	WIP	The App needs user to login, the program will open browser instance and wait for user to enter login credentials.
3	Scrape the content of the pages. Algo: beautifulSoup	Complete	
4	Data cleanup: Remove special characters, stopwords (Algo using NLTK)	WIP	
5	Use ranking algorithm BM25 to calculate rank of the relevant documents and select top 3.	Not yet started	
6	Show first 50 (or less) characters of the document in the UI, along with the link to the document.	Not yet started	
7	Write JS code to create a chrome extension to design UI for user to enter search query and show results.	WIP	