CS410 Fall 2020 Final Project Proposal: Improving Expert Search System

1. I am doing this project individually. My name is Mengyu Zhai and NetID is mzhai.

2. In this project, I choose to improve the Expert Search System (option 2.2), and more specifically, the subtopic of *Extracting relevant information from faculty bios.* I plan to add the function of extracting faculty research interests, so users can search faculty by their research interests.

3. Dataset to be used is the bios available under MP2.3 on Coursera Week 5: complied_dataset.zip. According to the introduction on the Coursera page, all the faculty bio submissions from MP2.1 that were submitted by Sep 20 were compiled into this dataset, and there are a total of 16492 unique faculty bios.

Algorithms or techniques to be used is *topic mining* on the bios available under MP2.3 on Coursera. The top-keywords per topic are the common research interests in Engineering and Science.

4. I am adding a function that is not in the original system, so it is relatively easy to demonstrate that it works as expected – In the final demo, I will perform some search with some research interests and show that it can really extract the relevant results.

5. I plan to utilize the system and code given, and then add & modify codes that can achieve the added function.

6. As the original code is in Python, I plan to use Python as the programming language.

7. As an individual project, its planned workload is about 20 hours, not including the time used for this proposal. I may spend more hours because I am not a Python expert, but I wouldn't count much of that into the workload of the project.

Main Tasks	Estimated Time Cost
Manually pick common research interest areas (topic terms) from a typical Engineering and Science department (plan to use UIUC CS department faculty's research interests pool https://cs.illinois.edu/about/people/all-faculty)	3 hours
Study original system code and system	2 hours
Work on the new function code and testing	13 hours
Make the final demo	1 hour
Upload all files (proposal, progress report, project code and documentation, project presentation/demo)	1 hour