1. Which tasks have been completed?

We have extracted source code from git hub <https://github.com/CS410Assignments/ExpertSearch>. Created local repository and troubleshot few issues and finally we were able to run expertSearch on localhost. We have started analyzing existing code and working on auto-crawler.

1. Which tasks are pending?

First, we are planning to create a python script to automatically crawl the faculty pages and extract all the bios data. We are using faculty webpage URL from MP 2.3 dataset for positive examples and will add some negative cases as well. From this bios data dump, we will also try to add the text content extracted from the faculty research interest page if available.

Next, we will work on the topic mining part. We will explore PLSA, LDA etc. algorithm to see which one is the best fit for our use case. We will work on the model creation manually in Jupyter Notebooks and will check and analyze the results to understand which model is better.

1. Are you facing any challenges?

We have faced few technical challenges while creating local repo. We are using Anaconda for our project due to some issues with Mac OS Big Sur’s built-in python versions. When we enabled python 2.7 environment in Anaconda, baseline version of the expertSearch didn’t work. There are few missing packages in the baseline library, and due to this we were not able to run expertSearch on localhost. we have troubleshot this issue and installed missing packages to fix this problem. We have faced few challenges with Gunicorn 19.10.0 as well and resolved these issues successfully, finally expertSearch is running on our local machine.

We have another challenge as well which is more critical. We were a 2-member team, and my teammate (Indranil Guha, Net ID: iguha4) went back to India due to a family medical emergency, so we have a resource crunch now; due to this we need to descope few items to adjust the workload. We will not be able to focus on UI integration part, also UI improvement part (which we marked as optional in the original proposal). We are only focusing on auto-crawler and topic-mining implementation as standalone python code, most likely in Jupyter Notebooks as mentioned above in #2.