CS 410 Project Progress Report

Course Topic Explorer (Free Topic)

Team: HREC
Members: Harita Reddy (haritar2) and Eric Crawford (ecraw3)

1) Which tasks have been completed?

The completed tasks are as follows:
- The backend setup is up and running.
- University websites' scraping for course information has been completed and the data is stored in pickle files to avoid frequent scraping.
- Implementation of search functions is completed and is returning the expected results for sample queries.
- The backend has been integrated with the search implementation.
- The work on implementing the recommendation functions (for recommending programs based on search history) has started and is in the initial phase where we are testing out different ways of vectorizing the text. Currently, we have pulled the programs from a single university.

2) Which tasks are pending?

The pending tasks are as follows:
- Add detail screen to the frontend that shows more information on the users selected course
- Optimize the speed of the backend api.
- Return the search results in a structured JSON format for the frontend.
- Complete a working recommender implementation.

3) Are you facing any challenges?

The challenges that we faced are:
- Meta does not work with web servers like Django and Flask. There is a known issue with deadlocks in the pybind11 library that meta depends on. Remedied this issue by using the Falcon framework instead after investigating similar issues reported by other users (Issue Link). If time permits, we may also look into using AWS lambda and API gateway.
- Text recommender function based on TF-IDF vectors is not returning the expected results for recommendation to the user. We need to further investigate how to improve the results returned by the recommender.

- Scraping different websites with common code was challenging because different websites have different HTML layouts. But as originally planned, we were able to scrape courses from 10 different universities which have an ideal layout.

4) Answers to reviews

Question: I was wondering what kinds of universities / programs you planned on including for your project (like schools with a wide variety of programs or smaller engineering or liberal arts schools).

We are including courses from universities with a wide variety of programs, like UIUC, Georgia Tech etc.

Suggestion: Include the % match of the program to the user's recent searches or just a bit of extra information like "because you've recently searched ___", <insert program> is a great fit for you because of their excellent computer science department!

That is a great suggestion and we will try to implement this if the time permits. However, we are not planning to include this in the milestones for the project since we already have a significant amount of work to be completed as part of the milestones. We will take this up as an optional task, and if we are not able to do this by the end of the project timeline, we will add the feature later.

Suggestion: Finish a minimum viable deliverable and then enhance it to your aspired functionality with the time you have left. Get something working quickly.
Suggestion: Time estimates seem to be a bit aggressive to achieve all the milestones.

We are implementing a minimum viable deliverable first, as suggested and then we will work on improving, such as improving the search or recommendation functionality further. We believe the minimum viable deliverable will be completed as per the timelines based on the progress we have made till now.