

CS 410 Project Progress Report

PS: We added two more tasks for our project:

Task 0: Setting up local development environment

Task 7: Setting up a docker environment and providing a public url to the users. (One of our teammates is especially interested in this part of work and will give a try.)

1) Which tasks have been completed?

Task 0: Set up a local development environment.

Status:

All the teammates have successfully set up a local development environment for the ExpertSearch project (<https://github.com/CS410Assignments/ExpertSearch>) and could implement expertise searches on the local website: <http://localhost:8095/>. We can now get a list of faculty information, including name, university, location, and a simple bio, using a specific expertise search. We can also visit the faculty's email and web page using the hyperlinks in the list. Every teammate is familiar with the ExpertSearch project and clear with our following work. In the future, we can always make some simple tests before we commit the changes.

2) Which tasks are pending?

Task 1~3: After showing the searching demo and discussing the specific methods on task 1~3, team members who are responsible for these tasks are now clear with the following work. Task 1 focuses on grabbing all possible links from a given link. Task 2 will choose faculty links from all the grabbed links using the Regex method. Task 3 focuses on identifying faculty links by binary classifier.

Task 4: What's pending of task 6 are transforming current codes into somewhat adaptable to what the team has, which being able to allow task 6 seamlessly blended in. And tuning major functions like EM, log-likelihood and PLSA as result it is versatile. Besides continually re-learning given lecture videos, finding topic modeling examples from open source would be helpful.

Task 5: Writing a topic mining function in extracting common search areas.

Status: Just discussed with the team to finalize the requirements. We would like to add more information into the web page like: **Research Interests ...**

Plan:

Will test the current framework to know how to return new fields into json

-- Nov 20, 2021

Implement new functions to obtain the new fields per faculty's home page

-- Nov 28, 2021

3) Are you facing any challenges?

Task 4: Topic modeling requires me to rewatch lecture videos more than twice. For anyone without a solid computer science background like myself. It's tough one even with my teammates' guidance. Challenges include the necessity of understanding a wide range of knowledge in order to work with teammates seamlessly, like LDA, Expectation-Maximum Algorithm, PLSA. Understanding these areas at every granularity helps substantially when bridging code with my teammates. After gaining solid understanding of what topic modeling

requires, being able to transform conceptual knowledge into codes is one of the most crucial parts of this task. But I'm in a great team with talented students. We continually grow and share, in hope to make our project standout and meaningful.

Task 7: To set up docker environment and provide a public url to the users (like TA)

Status:

Have spent 16 hours learning docker, gunicorn, nginx and web service.