

CS 410 Project Progress Report

Completed Tasks

The tasks listed below have been completed:

- Data preparation: cleaning, scraping for additional data, and enrichment

Pending Tasks

The tasks listed below are pending / in progress:

- Recommendation engine - started
- Frontend interface - pending
- Analysis and testing - pending
- Documentation and report - pending

Challenges

One challenge was with the dataset: there do not appear to be any datasets that exist for recommendation of coding projects, making it difficult to get started with a model and come up with an evaluation strategy. A dataset with just coding project records has since been found and the intent to use implicit and explicit feedback as evaluation, but a golden-standard dataset would have been nice. Another challenge was with data enrichment: initially, there was a small learning curve in fetching additional data to improve the recommendations by querying the GitHub API using GraphQL. The final / current challenge is modelling: this is a work-in-progress, however it was not immediately obvious how to apply content-based filtering based on the lecture content.

Review feedback

It was suggested to use implicit feedback in addition to explicit feedback for evaluation, this will be taken into detailed consideration when the project reaches that stage.

It was suggested to add more information about the dataset: it was available on Kaggle, scraped from GitHub and outlines: the owner, repository name, description, tags, last updated date, number of stars, language and URL of the top 980 most starred GitHub repositories/projects.