
Final Project - Proposal

Please read the final project description posted on iLearn or Google Drive (Assignments->FinalProject). Once you understand the final project description, then you can proceed to read the following description of the Final Project Proposal.

The Final Project can be done in groups (3 students). Your team composition is found on the Google Sheet https://docs.google.com/spreadsheets/d/1xCryV1JFCktUuJSsx00vNJQgVbwUs0JPKfzeoNj_HfE/edit?usp=sharing. You will note each team has a team-name. This team name is used to create your final project repository on Github.

I have already created the repositories for all the teams, and added the team members to the appropriate repository (if you provided your github repository). If you did not provide your github account on the Google Sheet, then you will need to 1) add your github account to the Google Sheet, and 2) add yourself to the appropriate repository. Lookup your 'team name' and ensure you are able to access your final project repository. Here is the link to the assignment <https://classroom.github.com/g/QLIXq9FQ>.

Since the final project provides each team some flexibility to define the parameters of their project (this was request based on past course offerings), we are asking that each team submit a short proposal describing what they plan to do for the final project within the project parameters. Each project proposal should have the following template.

- Team Name
- Team Members
- Team Github repo link
- Description of Part 1 - Crawling: Will you build a crawler or stream tweets? What language do you plan to use? How will you divide the work? Which type of pages do you plan to crawl (edu, gov, etc.) Have you thought about libraries that you plan to use to help you parse HTML if you build a crawler?
- Description of Part 2 - Retrieval: For this part you are asked to build an index/retrieval system using Lucene / Elastic Search. This portion should be similar whether you choose a Crawling or Tweet Collection. Did you read documentation provided by the TA on Lucene/Elastic Search?
- Description of Part 3 - Extension: What do you plan to do satisfy the third requirement of the project. Include as much detail as possible. We would like to read this portion to ensure it is sufficient for the final project requirement.

Submit 1 page proposal to iLearn. One submission per group is sufficient.