

CS410 Final Project Proposal: Free Topic – Topic Mining

David Tan Sang Tran (davidtt2@illinois.edu)

<https://github.com/davidtt2/CourseProject>

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members. One team member

| Name | NetID | Captain |
|---------------------|----------|---------|
| David Tan Sang Tran | davidtt2 | Yes |

2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

Free Topic/Task: I plan to create a website that displays information about top software companies, including technologies used, headquarters information, links to applications, internship opportunities, and other relevant information. (This project would be similar to the ExpertSearch, but I will build this project from scratch).

Interesting: This will be interesting because it would be good for Computer Science students to see a compiled list of all relevant companies that they can apply to after graduation.

Technologies, Tools, Systems: I plan to use Python with a text mining algorithm to parse links and collect data about all of the top software companies available. Selenium and requests will be used to collect the information, and then Pandas will be used to transfer the data into csv sheets. After the data is collected, TypeScript and Angular will be used to display the data and UI. Datasets will be the internet and various links which would be where the information is collected.

Expected Outcome: The expected outcome is a website that houses all the top technology company information for students to view and use for applications.

Self-Evaluation: I will evaluate my work based on how well I meet my own expectations. I expect a fully functioning website that will have a nice UI and all the needed information about the various companies.

3. Which programming language do you plan to use?

To collect the data, I will use Python, Selenium, Requests, and Pandas.

For the website, I will use Angular, TypeScript, ng-charts, and Angular Material library.

4. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

| Task | Estimated Hours |
|--|-----------------|
| Research and UI Mockup | 3 Hours |
| Topic Mining to Retrieve Data in Python | 10 Hours |
| Parsing Useful Information from Data Retrieved | 5 Hours |
| Developing the UI in Angular | 10 Hours |
| Connecting the Angular UI with the Data | 2 Hours |
| Testing the UI and Python | 5 Hours |
| Total: | 35 Hours |