CS 122 Project Proposal

Team Members: An Li, Nathan Lee, David Gonzalez-Velez

Description: We want to create a platform that aggregates data about various high school summer programs across the United States. Users will be able to go to this database and use it as a search engine. They would be able to define their search by criteria such as program cost, dates/duration, deadlines, location, field of interest, and year in high school. The goal of this website is to improve the accessibility of summer program information to all high school students.

In the summer program database sphere, there exist a handful of search engines that tailor summer programs based on inputted personal interests. Summer Discovery, Teen Life, and USummer are examples of competing websites in this market, however, none of them are comprehensive nor easy to navigate. While each site generally informs their audience of summer programs which correspond to indicated interest, they lack developed infrastructure and visual appeal. In order to streamline the summer-program search for consumers, the website must smoothly facilitate the transition of the inputted information from the consumer and deliver a desirable product (a customized and thorough display of summer programs).

Data Sources:
We plan to scrape the following in this order (time permitting):
- Existing databases that have limited search features and programs, such as Summer Discovery, Teen Life, USummer, Pathway to Science, etc.
- University websites that offer high school programs
- Companies with international high school programs
    - Rustic Pathways
    - National Geographic Student Expeditions
    - Bold Earth
    - Discover Corps
    - Where There Be Dragons
    - GLA: Global Leadership Adventures
    - Projects Abroad
    - Go Abroad
    - Putney Student Travel
    - ACIS (not sure if just summer): American Council for International Studies
    - Teen Missions International
    - VISIONS Service Adventures
    - The Experiment In International Living
    - Go Overseas
    - GoEco
    - GVI
    - CIEE
    - Mitzvah Corps

General Tasks and Timeline:
- Design generic crawler for existing databases summer programs
    - Week 6: We plan to have at least one working version for 2-3 databases

- Aggregate data (i.e. cost, duration, deadline, etc.) from the web pages into different queries
  - Week 7: We would like to have the ability to store and then combine the information from different queries as well
- Create platform or webpage for interaction of such data
  - Week 8: We would like to have a generic shell by this time and hopefully be implementing a design as well
- Condense and clean data in order to filter programs into different categories
  - Week 9: We hope to be able to convert the data into more usable forms for presentation
- Expand scope of our data sources by incorporating other websites if time permitting
  - Week 10: We would also be using this time to fix and adjust our previous steps if we were too ambitious