Liam Juskevice

Progress Report

Unfortunately this project is still in the preliminary stages as far as my work goes, and I will definitely need to pick up the pace to reach my target 20 hours and my overall goal for the project. However, I will share what that preliminary work entails. After looking at the reviews I’ve decided to put most of my focus on the web crawling and back end at the risk of likely having a rather bare bones front end at the end. I feel the former two will end up taking close to 20 hours on their own. Well, if you consider how long the actual act of web crawling can take it could easily exceed that. Regardless, I have decided to use a MySQL database to store and structure the text data I gather about every fish.

The database is structured rather simply at the moment. The primary table is naturally the fish table, where every row is a species of fish and each column is a relevant piece of information about them such as their size, temperature range, and considerably more. Besides that I have decided to create a separate table for countries and regions because I want to track the geographic origins of each fish species. Since fish can be from multiple countries, and less frequently different regions I’ll track the fish origins using the fishCountries and fishRegions relationship table. Note that the region table is essentially just the seven continents minus Antarctica because I’m pretty sure the fish trade doesn’t have any freshwater fish from that region. Any ways, my immediate next step would simply be to populate these tables with every currently recognized country on earth and assign a region to each one of them. This step probably won’t take too long at all. Next comes the web crawling.

For the web crawling I have decided to use two data sources for now, SeriouslyFish.com and LiveAquaria.com. The former attempts to be a comprehensive database on non-marine aquarium fish and I would likely have every fish species I could realistically want to use included after crawling through their entire collection. Crawling LiveAquaria is a less necessary step that can provide some supporting information about a smaller subset of those fish species. For instance a more succinct summary of their temperament (it’s pretty important to not pair peaceful fish with aggressive fish after all). I just need to design the crawler to go through everything SeriouslyFish and potentially LiveAquaria have to offer. After I do that I’ll need to take stock of how much time I’ve spent and how much time I have remaining. This will inform my decision on how sophisticated or skeletal the front end aspect of this project will be.