Final Project Proposal:

Understanding the University of Michigan Museum of Art Collection

By Alex Carey

Overview

For my final 507 project, I will build a program that will scrape and crawl the University of Michigan Museum of Art online collection platform, The UMMA Exchange. This website has information about every piece of artwork in the UMMA's collection, including each piece's title, artist, creation date, medium, and other descriptive information about each item. Scraping and crawling the many pages of this website will achieve the total *challenge score* of eight.

The Program and Tools

My program will ask the user to enter a search term, which I will then append to a baseurl to create a new url that will generate a search query. The program will return a numbered list of results while simultaneously creating a database of artwork from that search query. If a user then chooses "graph" as a command, Ploty will open a pie chart visualizing the different artists represented in a query if a search term is a keyword, or the different medium represented if the search term is an artist's name. The user can also enter a number that corresponds to a given work of art in the returned list of results, and an Exchange page about that particular artwork will open in a web browser, like this. At any time, the user can exit the program and start a new search. Additionally, the program will implement caching and correspond with a series of Unit Tests to ensure the program works as expected.

The Database

The database, to be named umma.sqlite, will have two tables, one called Art with general information about all the artwork that is generated by each query, and a second called Artists with information about each artist who appears in the query results. This table will include the artist's name, birthdate, nationality, etc, and the data will be gathered using the MediaWiki API. Each piece of art in the Art table will have an ArtistKey that aligns with an artist's auto-incremented Id in the Artist table.