Design Diary

One particular struggle I encountered was overcoming a certificate error. On Feb 19 I had an error when testing the introductory code that read SSI.SSLCertVerificationError: local issuer certificate (\_SSI.C:1108). After asking the instructor for help, the issue was overcome by searching the problem on the web, and using a recommended pip install. After the pip install, the introductory code could run without errors.

Advice to help a future student succeed on this assignment would be to research web scraping, beautiful soup, python syntax, and html. The assignment not only requires python knowledge but some knowledge of html for the web scraping aspect. Mainly, future students need to know how to work with divs, hrefs, and other aspects when attempting to grab links or titles off of a website. Likewise, beautiful soup has aspects to it that can be helpful in the assignment, and as always python syntax is required for writing the code.

The most fun aspect of the assignment was coming up with the idea itself, and designing the code to that idea. Mainly, grabbing links and titles of a certain search term was the idea, and it was fun to succeed in this aspect. The search terms were placed in a list that can then be displayed.

The most challenging aspect of the assignment was learning how to use the html tags in relation to the python code. It was not immediately apparent which html tags related to which links and titles to the uninitiated. Upon using an inspect element command on the webpage, there exists an extensive collection of html content. This can be intimidating to sift through when just learning html.

The most difficult aspect of the assignment was working out the final errors from the code. Many times, there exists nuance to the python syntax and ideas that can make or break the code. There were difficulties with combining the URLs into a single link, as the initial link was just the end of the URL. This required combining the two links together, applying a string of the main website to the second half of the URL. Another difficult aspect was removing certain links from the page that were not the desired results before storing into the list. While a page will have the links from a search result, it also includes links to similar articles, recently viewed pages, suggestions, contact pages, etc. These extra links included the ? character, so it was solved by searching the links for a ? and removing them from the list.

Suggestions for improving the assignment in the future might be extra instruction on html for those who have not used it much. Particularly on the topics of div and href, as those are most pertinent to locating links.