**CSCE 5200 - Information Retrieval and Web Search**

**Final Project**

**Harsha Buddana - 11521994**

**Aditya Ananthula - 11560316**

**Vamshi Peddi - 11505469**

**Sairam Reddy Jillala - 11610556**

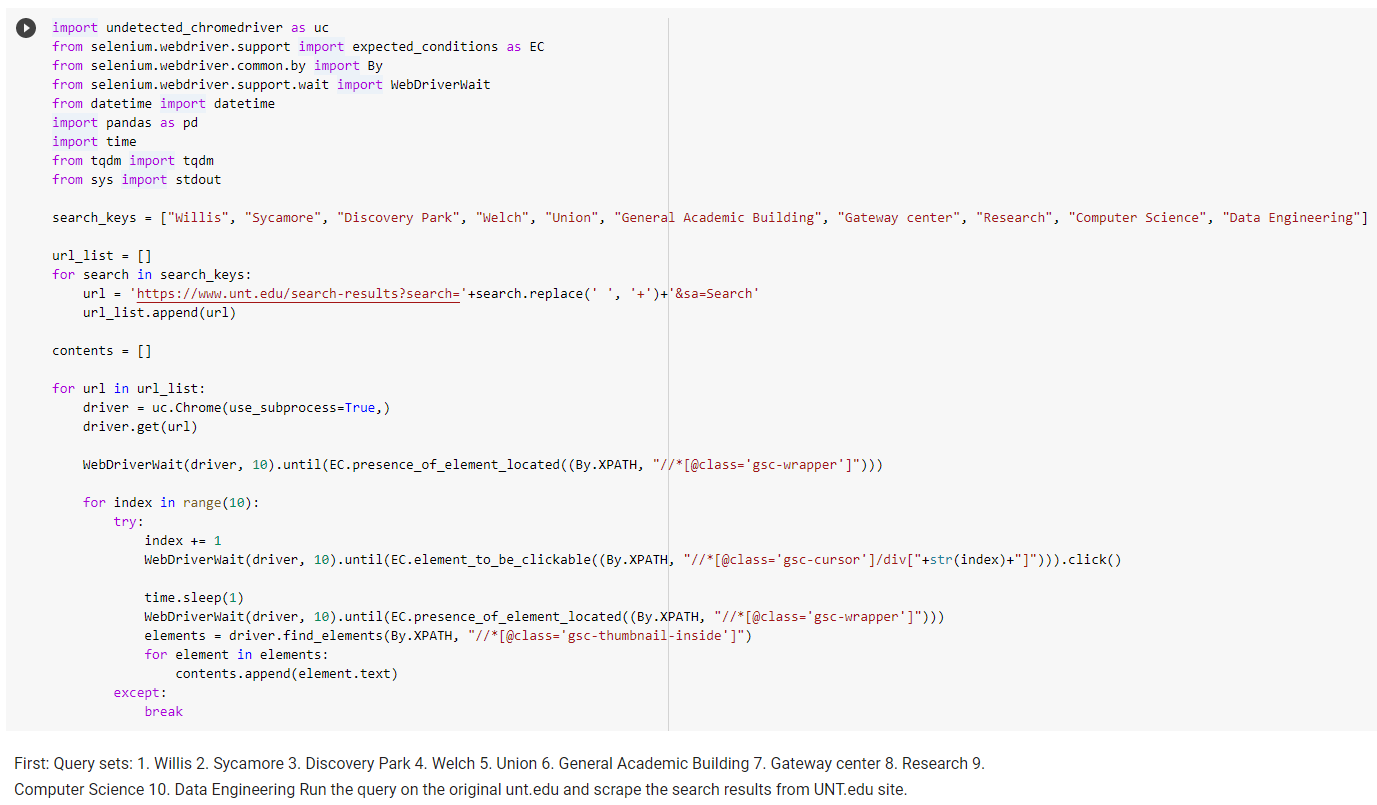
**Revanth Gurram -11520311**

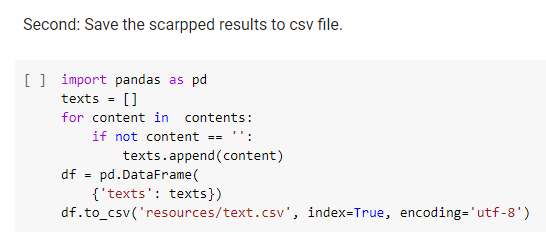
**Srikanth Gopi - 11514330**

The complete code is in an ipynb(Python Notebook). This file is in the zip file that we uploaded to Canvas. There are no explicit rules to run the program, just open the code in vscode and run the program.

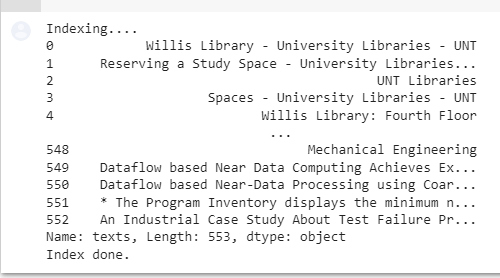
**Framework of the project**

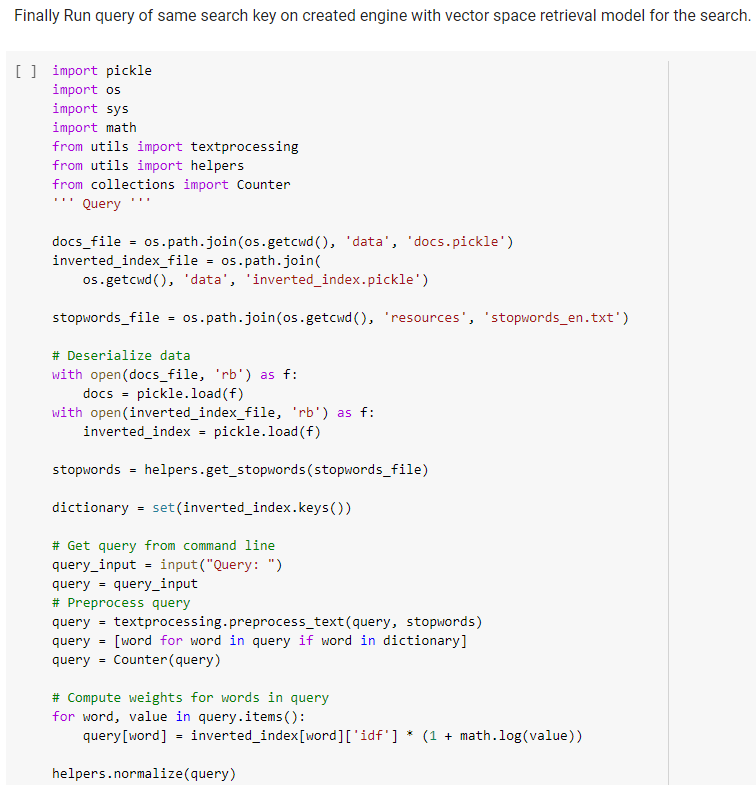
1. Crawled the website for a set of focused words.
2. Scraped the search results regarding the key words from the website.
3. Saved the results to a csv file.
4. Create a dictionary of words (Indexing) for the vector space retrieval model for the search and then create a search engine with a vector space retrieval model for the search.
5. Finally Run query for a search key on created engine with vector space retrieval model for the search.

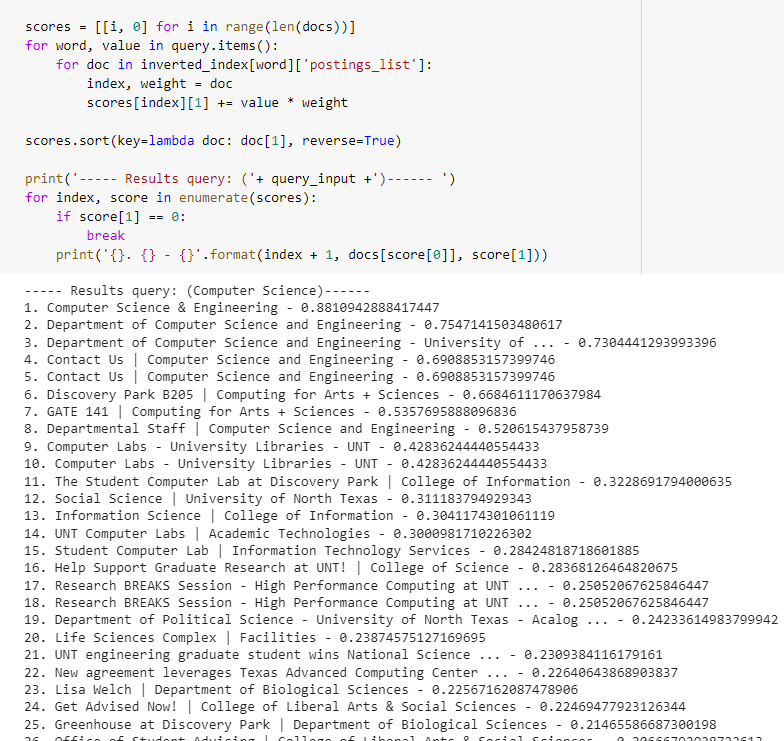


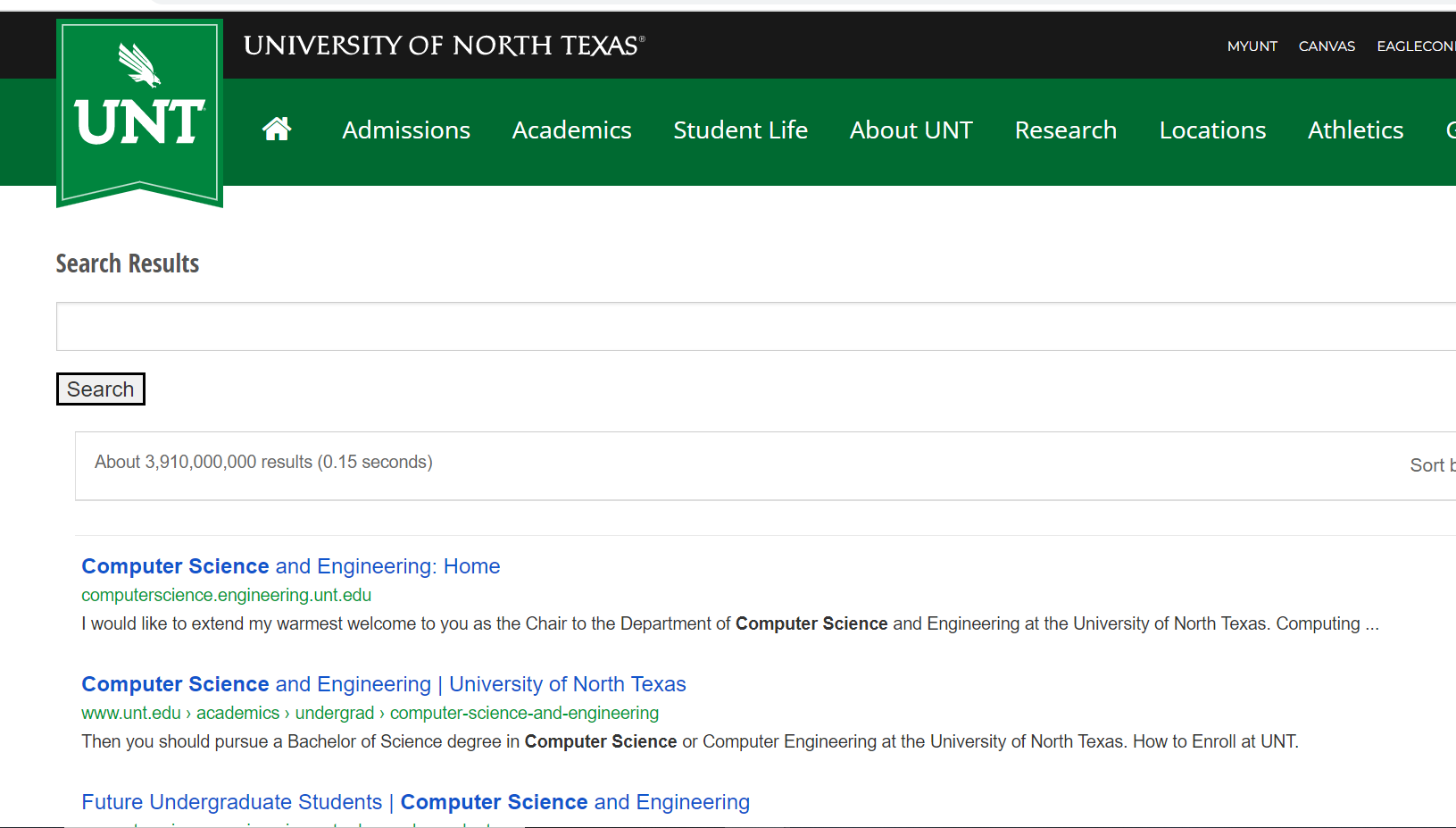


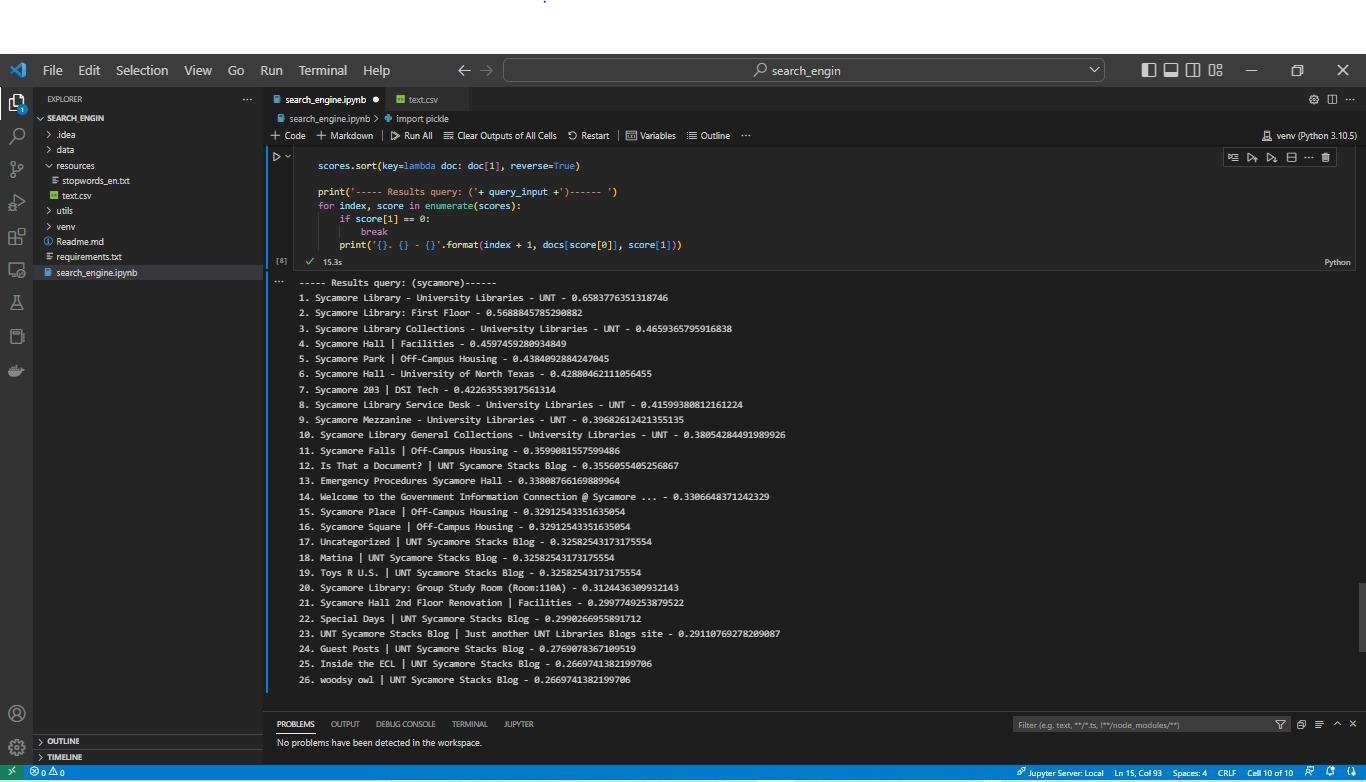


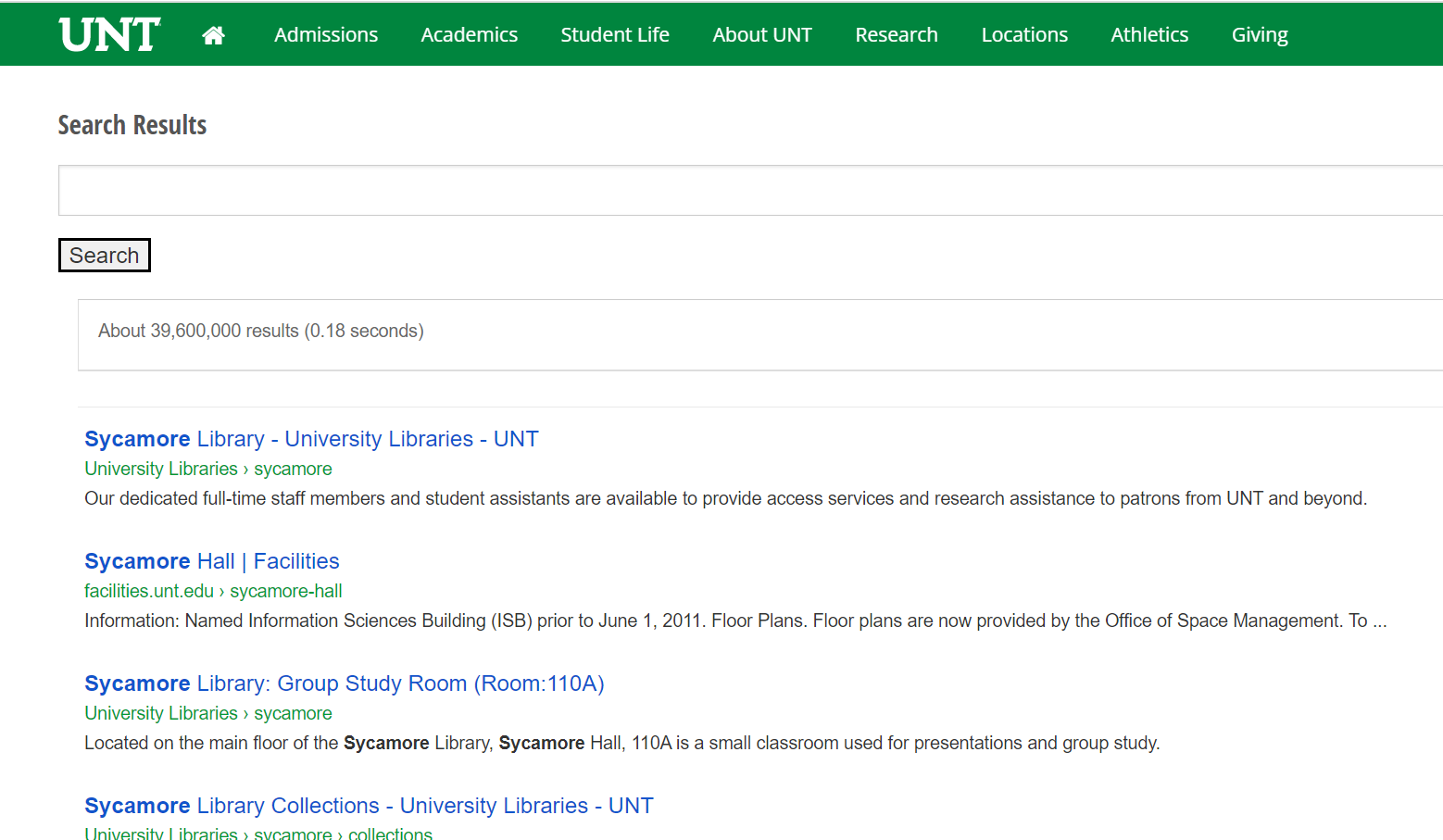


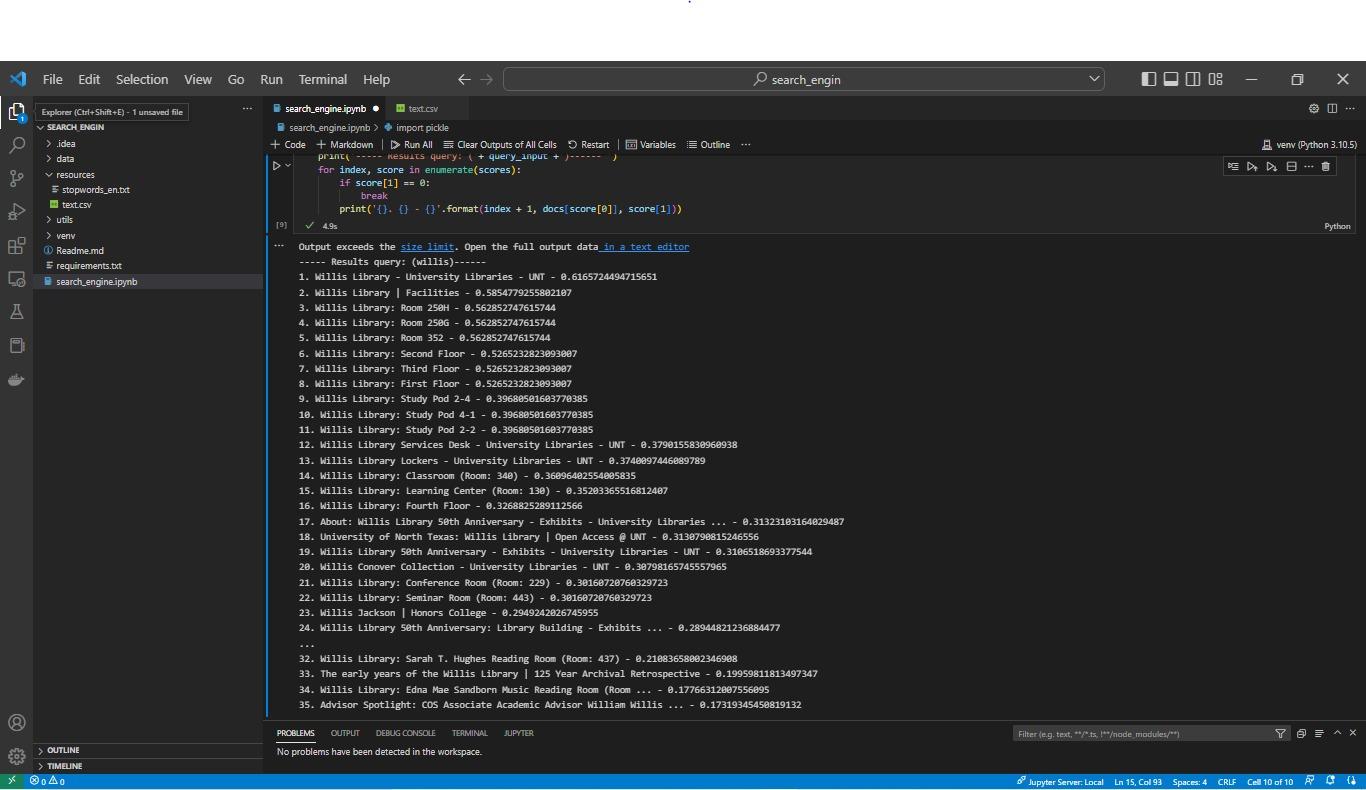


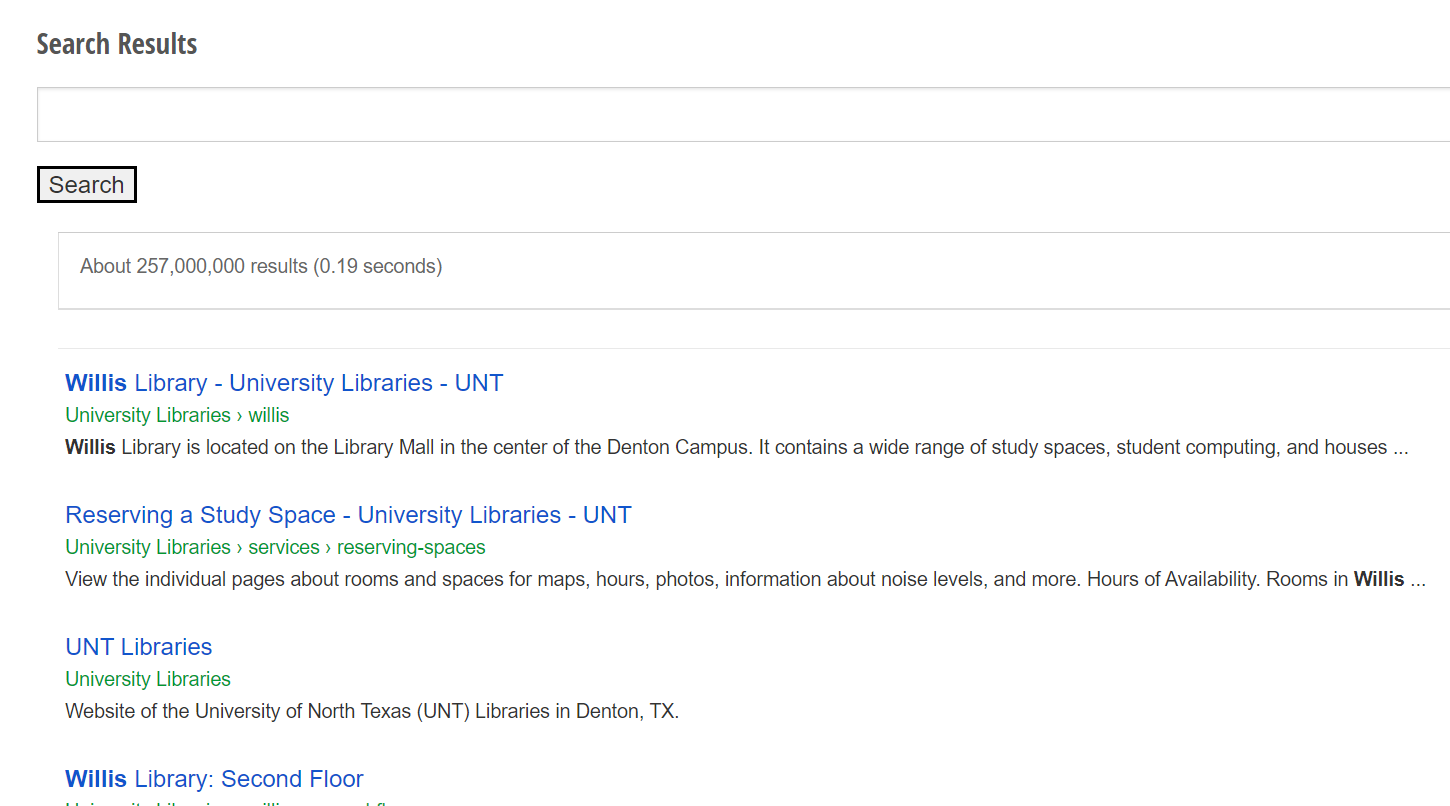












I have compared by entering the keywords both on the Original website and the search engine we have created. We can see that though our search engine doesn't give the exact results as the original. But it still tries to fetch the relevant terms from the dictionary.

**Challenges Faced** -

1. Chrome version compatibility issues.
2. Problem of Captcha and stopping the scraping.