**Search Engine**

Implement the simplified Search Engine described in Section 23.5.4 for the pages of a small Web site. Use all the words in the pages of the site as index terms, excluding stop words such as articles, prepositions, and pronouns.

## dependencies

1. python
2. beatifulsoup (pip install beautifulsoup4)
3. nltk (pip install nltk)
4. pip install requests

I created my project in Python because it is a great language for creating dynamic objects and easy to code. I output the occurrence list and the trie for each, so you can examine the inputs and methodology.

I used beautifulsoup to extract webpages. I then stripped all html and css properties and stored the text. I then open data and process the text. I remove all stop words and punctuation and split the words into an array.

From here I count the number of times a word occurs in the list and create an occurrence list. I then take every word from this occurrence list and pass it into a function I made to create a standard trie which has the index of its location in the occurrence list, as outlined by Section 23.5.4. I also created a function that searches through the tries to find a word and returns the index of that search word in the occurrence list if it exists.

Boundary Conditions:

1. Half words will not be searched, i.e., prefix search will not be executed. Only

complete words will be searched.

2. Numbers and special characters will be neglected from the webpages as well

as from the user input.