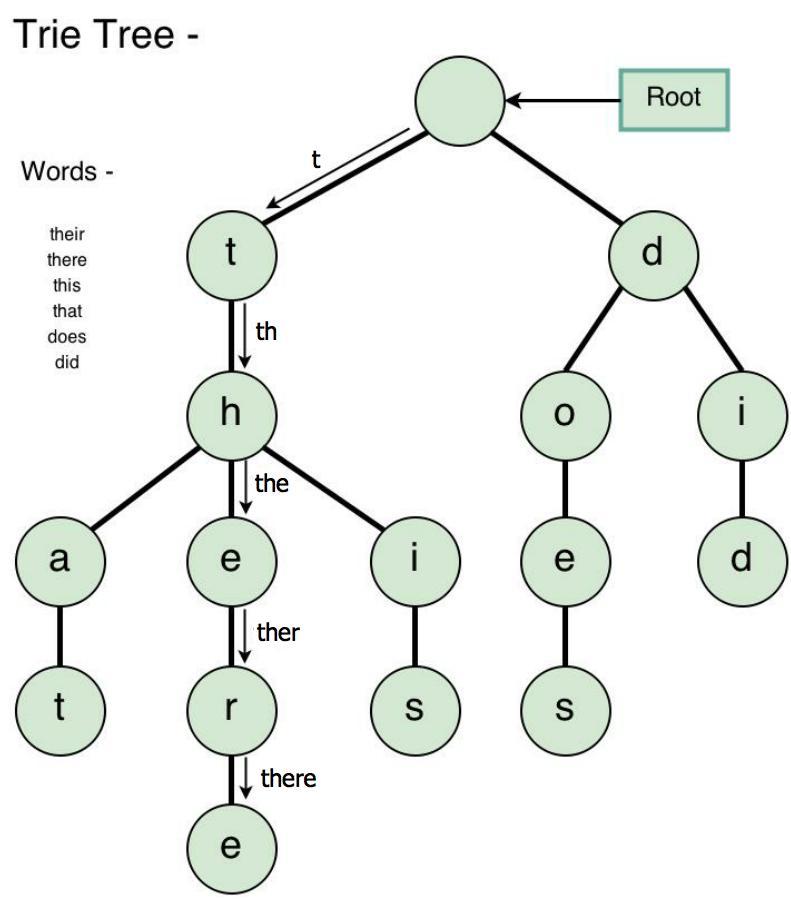
CS 600 Search Engine

Atiq Patel

I have used several links that are present in the “Input.txt” file. I have coded this project using Python and the data structure I have used is Trie which works as shown in the figure.



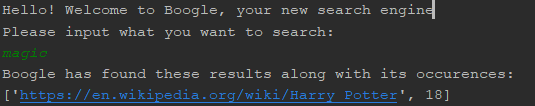
I have scraped the websites using BeautifulSoup and all the data is stored in Trie. Stop Words have been stored in a “StopWords.txt” file. We have created a StopWords Trie as well. This is being utilized when we are searching for words from the links and then we compare each words to the stopwords. If it matches then it throws an (“Please don’t enter StopWords”) error.

This engine throws the link and its occurrence which occurs the highest among all the links. This is achieved by using a Page Comparator which compares the occurrences of the given word, and then searches it among all the other links if present inside them and gets the one with the most occurred link and its count.

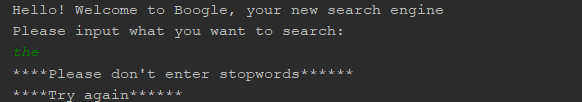
I have several different python files that are needed to execute the project. To start the project all you need to do is run the “SE.py” in the same directory as the other files as well as change the path to your present path in “StopWords.py” for “Input.txt” and “StopWords.txt”.

The search engine “Boogle” asks you the phrase which you want to find, it will throw out the particular word or phrase if present among any links, along with its highest occurrence among the given links.

**Outputs:**



Example of a word being found in the website.



Example of an occurrence of a stopwords

My search engine can also find MULTIPLE KEYWORDS. Example below

