For this project is used links to read from the webpages using jsoup.

Reading paragraph from the webpages I extracted words and stored them in a trie data structure,

The find the word entered by the user, trie data structure is used to search as follows :



To find the count of occurences (ranking) of word from each webpage I used a hashmap for each link which displays their word count respectively. If the word is not present in one of the document the count is shown as zero.

To remove the stop words, as we read each word from the webpage’s paragraph pass that word to the isStopword function containing a list of all stopwords. If a match is found that word is not inserted into a trie datastructure as well the hashmap.

Also, to display the total of all the words from all the documents I maintained a separate map which displays the words as well their total occurences found from all the documents.