

# **CS 146 Programing Assignment 2**

## **Google Search Engine Results Page Simulator**

### **Using Quick Sort, BST and Bucket Sort**

Yu Xiu

Advisor: Dr. Mike Wu

Nov. 16, 2018

## TABLE OF CONTENTS

<b>CS 146 Programing Assignment 2 .....</b>	<b>1</b>
<b>Illustration of this Google Search Engine Simulator .....</b>	<b>3</b>
<b>Design .....</b>	<b>3</b>
<b>A List of classes and function calls and Explanations of each purpose.....</b>	<b>4</b>
<b>Class 1: GoogleSinmulator .....</b>	<b>4</b>
Functions and Explanation: .....	4
<b>Class 2: MyCrawler .....</b>	<b>4</b>
Functions and Explanation: .....	5
<b>Class 3: WebPageInformation .....</b>	<b>5</b>
Functions and Explanation: .....	5
<b>Class 4: HeapSort .....</b>	<b>6</b>
Functions and Explanation: .....	6
<b>Class 5: KeywordInformation .....</b>	<b>7</b>
Functions Explanation: .....	7
<b>Class 6: QuickSort .....</b>	<b>8</b>
Functions Explanation: .....	8
<b>Class 7: BucketSort .....</b>	<b>8</b>
Functions Explanation: .....	8
<b>Class 8: BinarySearchTree .....</b>	<b>9</b>
Classes and Functions Explanation: .....	9
<b>Screen shots of each simulation process.....</b>	<b>9</b>
<b>Screen Shots.....</b>	<b>9</b>
Input: corgi.....	10
Input: sugar .....	10
Input: cat .....	10
Input: tea .....	11
Input: rock.....	11
<b>Test BST Search: .....</b>	<b>12</b>
<b>Test BST Deletion: .....</b>	<b>13</b>
<b>Test BST insertion:.....</b>	<b>14</b>
<b>Test Bucket sort: Sorting company names: .....</b>	<b>15</b>
<b>The procedure of how to unzip files, install application, and run codes .....</b>	<b>15</b>
<b>PA1-Yu-Xiu.zip .....</b>	<b>15</b>
<b>Run the jar file in console:.....</b>	<b>16</b>
<b>Procedure of how to run my codes: .....</b>	<b>16</b>
<b>Problems encountered during the implementation.....</b>	<b>23</b>
<b>Lessons Learned.....</b>	<b>24</b>

## ILLUSTRATION OF THIS GOOGLE SEARCH ENGINE SIMULATOR

In this Google Search Engine Simulator project, the purpose is to simulate the ranking method of displaying the results of keyword based searching. To search web pages based on users input, we need a web crawler to crawl web pages based on the input. In order to simulate Google Search Engine, it requires a design of how to display the searching results to the users. In this project, the ranking of web pages is primarily based on a calculation of four factors. The first factor is the **frequency of keywords** within the web page, which gives the number of recurrence of the keyword in text of a crawled web page. The second factor is the **existence length** of the web page, which shows how long the web pages had existed. The third factor is the number of other web pages that **link** to this web page, and the last factor is how much the web page pay for its **advertisement**. There are **four weights**,  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$ , which can be changed by user. The total score of each web page comes from the formula:

$$\text{Total Score} = a_1 * \text{keyword frequency} + a_2 * \text{web page exist age} + a_3 * \text{number of outgoing links} + a_4 * \text{advertisement payment}$$

We will have scores calculated from each web page. By developing a simple Web Crawler in Java, we can collect information from web pages and find the four factor scores. Each web page has a total score based on the formula provided.

After calculating and comparing scores of each page, we will use **quick sort** algorithm to sort those scores and display the ranks in a descending order, which means the user will see the searching results with the highest score ranked on the top. In this project, the quick sort would sort 30 URLs with their web page information that we retrieved and saved from web crawler. Based on the scores or page ranks of 30 URLs, the **Binary Search Tree** would be used to search a page rank, insert a URL, delete a URL, and sort the information with inorder tree walk. In demand to avoid each time when the user input a keyword the program would test BST, there is an option provided for user whether testing BST features. Finally, this project also implement **bucket sort** to sort the top first popular key's URLs and information. The total time complexity is  $O(n \lg n)$ .

### DESIGN:

This project provides command line interface. Initially, there is a chance for the user to input the four coefficients or weights,  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$  for the total score formula that later would be used to calculate the rank of the web page. Then the user would input a keyword. After that, the web crawler start to crawl web pages based on the keyword and display the top 30 URLs. Then the user would be asked whether test BST, if Yes, the user

could choose from 5 options, which allows user to insert, delete URLs or search a specific page rank, and one can also choose to exit the BST features. If the user chooses No, the user would not test BST, and keep entering keywords or exit the whole program. The details of how to run the code would be provided in the section of “procedure of how to run my code”.

## **A LIST OF CLASSES AND FUNCTION CALLS AND EXPLANATIONS OF EACH PURPOSE**

### **CLASS 1: GOOGLESINMULATOR**

**GoogleSimulator.java:** This class contains 3 functions: `main()` function, `crawlWebPages()`, and `testProcessBST()`. In this class, it configures and calls `MyCrawler` class. It reads user’s input and calls `quicksort()`, `bucketSort()`, `treeDelete()`, `treeSearch()`, and `treeInsert()`. And also, it displays the outputs of ranking, searching, deleting, inserting results and options.

#### **FUNCTIONS AND EXPLANATION:**

- 
1. **`crawlWebPages()`:** This function is from open source web crawler. It helps to develop a simple web crawler.  
Giving a seed for the controller, we will get the web pages that crawled based on user’s input.
  2. **`main(String[] args)`:** This function is the main function. It crawls the web pages and display the weights `a1`, `a2`, `a3`, `a4` for user to change. It calculates scores for every page and sort web pages by score then display them. It displays commands and results let users to read. It calls `quickSort()`, `bucketSort()`, and `testProcessBST()`. Time complexity is  $O(n \lg n)$ .
  3. **`testProcessBST(WebPageInformation[] webpages)`:** This function is designed out of the main while loop, which used to let user to repeatedly to input keyword. To do so, it can be avoided to put the searching page rank number, deleting URL, or inserting URL. It provides an option to let user to choose which feature they would want to test. For example: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit. When user enter a certain number from 0 to 4, the program would response the specific implementation of BST. When enter 0, it exits the test of BST; when enter 1, searching a node based on rand that user inputs; when enter 2, inserting a node based on user input; when enter 3, delete a node from tree; when enter 4, displaying sorted list with inorder tree walk method.

### **CLASS 2: MYCRAWLER**

**MyCrawler.java:** Using the link provided in the references from Dr. Wu, source No.2, a simple Web Crawler instance, crawler4j, is built in the program to collect web pages' information. It contains shouldVisit() and visit methods which give us a chance to save the information from each web page that we might use later.

#### FUNCTIONS AND EXPLANATION:

- 
1. **shouldVisit(Page referringPage, WebURL url):** Github writer uses this to filter URL.
  2. **visit(Page page):** Visit each web page and gives us a chance to save the information of each web pages.

#### CLASS 3: WEBPAGEINFORMATION

**WebPageInformation.java:** Each web page information would be store in this class after crawled and calculates scores of each web page. Using setters and getters to set and get web page information crawled from Web Crawler.

#### FUNCTIONS AND EXPLANATION:

- 
1. **compareTo(WebPageInformation i):** Compare the scores of two web pages.
  2. **toString():** A string representation of the results. Converting web pages information, which associated with the keyword into a string, and it would be printed out in main function in GoogleSimulator class.
  3. **CalculateScore(String keyword, double a1, double a2, double a3, double a4):** It calculates key word frequency by counting the occurrence out of a page's text.
  4. **getCompayName():** Splitting URL into string by cutting out the string before and after ".". For example, [www.google.com](http://www.google.com) -> google.
  5. **get\_CalculateScore():** get the calculated score.
  6. **getUrl():** get URL of a web page.
  7. **getText():** get text of a web page.
  8. **getHtml():** get html of a web page.
  9. **getOutGoingOutLink():** get out going links that points to the web page.
  10. **setUrl(String newUrl):** set a new URL for this web page.
  11. **setHtml(String newHtml):** set a new html for this web page.
  12. **setOutGoingLink(int OutGoinflink):** set a number of outgoing links.
  13. **setAge(int newAge):** set the age of a web page.
  14. **setRelevance(int newRelevance):** set number of relevant links link to the page.
  15. **setAds(int newAds):** set how much the company pay for the advertisements.

## CLASS 4: HEAPSORT

**HeapSort.java:** This class was used in PA1, and in this PA2, it only used to sort the top 10 most popular key word, which would be used to sort company names by using bucket sort. This class implements the heapsort sorting algorithm, which is an efficient algorithm to sort an array and supports the priority queue. It contains 10 functions: Parent, Left, Right, MaxHeapify, BuildMaxHeap, Heapsort, MaxHeapInsert, HeapExtractMax, HeapIncreaseKey and HeapMaximum methods to implement the sorting algorithm by sorting the scores obtained from each web page.

### FUNCTIONS AND EXPLANATION:

-----  
**1. Parent(int i):** This function defines the parent node by dividing the child index  $i$  by 2, and it will automatically taking floor of  $i / 2$  in java Time complexity is  $O(1)$ .

**2. Left(int i):** This function defines the left child by multiplying parent index by 2. Time complexity is  $O(1)$ .

**3. Right(int i):** This function defines the child node is found by multiplying parent index by 2. Time complexity is  $O(1)$ .

**4. MaxHeapify(E[] A, int i):** This function makes the largest value stored in the root and maintains the max-heap property. It lets the value at  $A[i]$  “float down” in the max-heap so that the subtree rooted at index  $i$  obeys the max-heap property. The array  $A$  takes a generic type, which allows us to sort different types later. In this function, it compares  $A[i]$  with its left child  $A[l]$  and right child  $A[r]$ , if  $A[i]$  is the largest value, then the subtree rooted at node  $i$  obeys the max-heapify. Otherwise, one of the two children have the largest element, and  $A[i]$  swap with  $A[largest]$ . We call Max-Heapify recursively on the subtree. Time complexity is  $O(\lg n)$ .

**5. BuildMaxHeap(E[] A, int i):** It goes through the remaining nodes of the tree and runs MaxHeapify for each of them and produces a max-heap from an unsorted input array. For example, we have 10-element input array  $A$  which is starting count at 1, firstly, index  $i$  refers to node index of 5 before the call of  $\text{MaxHeapify}(A, i)$ , the loop index  $i$  for the next iteration refers to node 4. Keep following this loop until down to index of 1 in the BuildMaxHeap function, then the MaxHeapify is called on a node, the two subtrees of the node are both max-heaps. Time complexity is  $O(n \lg n)$ .

**6. Heapsort(E[] A):** This function sorts the array of calculated scores of each web page in place. Firstly, the Heapsort starts by using BuildMaxHeap to build a max-heap on the input array  $A[0..n-1]$ , where  $n = A.length$ . In each iteration of the for loop, the index

i refers to the node at the final position of the array, since the maximum element of the array is stored at the root A[0], we can put it into the final position by swapping A[0] with A[i], where  $i = A.length - 1$ . If we discard the node A[i], we will have the heapSize decrements by 1. After swapping A[i] with A[0], we still need to keep the heap obeys the max-heap property, so that we call Maxheapify to do so. Time complexity is  $O(n \lg n)$ .

### Max-priority Queues:

**7. MaxHeapInsert(E[] A, E key):** Max-Heap-Insert function allows heap data structure to implement a priority queue. Insert an element to the heap tree. Time complexity is  $O(\lg n)$ .

**8. HeapExtractMax(E[] A):** It removes and returns the largest element of the max-heap. Firstly, it checks that if there exist an element to be extracted. If the heap is empty, there is an error showed up to the screen. If there exist element, it swaps the root A[0], which is the largest element in the heap, to A[heapSize - 1], which is the final element in the heap each of the time. Then, the heapSize decrements by 1. Calling the function MaxHeapify to maintain the max-heap priority for the rests of the nodes. After this procedure, we have the largest element extracted from the heap. Time complexity is  $O(\lg n)$ .

**9. HeapIncreaseKey(E[] A, int i, E key):** Increase the value of an element's key to a new value k, which is assumed to be at least as large as the current k. The procedure first updates the key of element A[i] to the new value. A[i] is the element that we want to increase its key. If  $A[i] > key$ , which A[i] is comparing with, there is an error. If  $A[i] = key$ , then when  $i > 1$  and  $A[Parent(i)] < A[i]$ , we swap A[Parent(i)] with A[i]. Time complexity of HeapIncreaseKey is  $O(\lg n)$ .

**10. HeapMaximum(E[] A):** This function returns the largest element A[0] in the heap. Time complexity is  $O(1)$ .

## CLASS 5: KEYWORDINFORMATION

**KeywordInformation.java:** This class stores the keyword and count number in a pair. The count numbers will be sorted using heapsort.

### FUNCTIONS EXPLANATION:

-----  
**1. KeywordInformation(String keyword, int count):** Constructor.

2. **compareTo(KeywordInformation other):** Comparing the two keywords' information, if `this.count > other.count`, return positive, if less than, return negative, and if they are equal to each other, return 0.
3. **toString():** Represent information of the object as a string. Converting keyword and its count into a string.

## CLASS 6: QUICKSORT

**QuickSort.java:** This class contains 2 functions, `partition(E[] A, int p, int r)` and `quickSort(E[] A, int p, int r)`. It implements the quick sorting algorithm, which would sort the 30 URLs and information, index, pageRank, total score and URL of the keyword.

### FUNCTIONS EXPLANATION:

- 
1. **partition(E[] A, int p, int r):** This function partition the array of the web page information by choosing pivot and comparing elements with the pivot. Comparing two web page information based on their total scores, and put the pivot in the right place.
  2. **quickSort(E[] A, int p, int r):** This function recursively called itself to partition the left and right subarrays.

## CLASS 7: BUCKETSORT

**BucketSort.java:** This class contains 2 functions, `bucketSort(ArrayList<WebPageInformation> A)` and `insertionSort(ArrayList<WebPageInformation> A)`. It implements the bucket sorting algorithm, which would sort the 30 company names, which are the split-string out of the domain of the URLs.

### FUNCTIONS EXPLANATION:

- 
1. **bucketSort(ArrayList<WebPageInformation> A, int p, int r):** In this function, an empty bucket B with arrayList typed web page information. It takes ArrayList A and inserts `A.get(i)` into different buckets of B, and later, it calls `insertionSort` function to sort the web page information's String of the company names (since the string of company names stored in `WebPageInformation` class) in the bucket. It also contains a `A.clear()` function, which clears A and puts the sorted company names from bucket B to A.



2. **insertionSort(ArrayList<WebPageInformation> A):** This function sorts the company names got from webPageInformation in the buckets of array B by comparing adjacent company names in a bucket in alphabetical order.

## CLASS 8: BINARYSEARCHTREE

**BinarySearchTree.java:** This class contains a node class and a tree class, and other 6 functions. In this class, it implements the binary search tree features and functions including search a node, insert a node, delete a node and sort all the nodes in inOrder tree walk method. For this project, it is going to implement the function of searching a pageRank, inserting URL, deleting URL and sort web page information in inOrder traversal.

### CLASSES AND FUNCTIONS EXPLANATION:

- 
1. **Class Node:** In this class, it defines the nodes of a tree including left reference, right reference, key, parent node, and WebPageInformation typed info. It contains 2 functions, a constructor Node (int PageRank, WebPageInformation info) and a toString(), which would be used when printing the node info.
  2. **Class tree:** Defining a root node.
  3. **BinarySearchTree():** default constructor.
  4. **treeSearch(Node x, int k):** Searching a PageRank in the tree, if the rank is smaller than node x, goes to x's left subtree, otherwise goes to its right subtree.
  5. **inorderTreeWalk(Node x):** Sorting nodes/ ranks in inorder traversal.
  6. **treeInsert(Tree T, Node Z):** Inserting an URL into the tree by comparing rank and find its right place to insert.
  7. **transplant(Tree T, Node u, Node v):** Transplanting subtrees to another subtree, it would be called in treeDelete function.
  8. **treeDelete(Tree T, Node z):** Deleting a URL from tree. It handles 3 cases. The node z has no children; it has one child, either left or right; and it has two children. If no children, directly delete node z. If z has no left child, transplant right subtree of z. If z has no right child, transplant left subtree. If z has two children, find successor and transplant the correct subtree.

## SCREEN SHOTS OF EACH SIMULATION PROCESS

### SCREEN SHOTS:

a1 = 5, a2 = 2, a3 = 3, a4 = 1

INPUT: CORGI.

Output: top 30 search results with pageRank, index, total score, four factors, company names, and URL:

```
PageRank 1 Idx 2 Total Score = 1100.0 Four Factors = { 168, 23, 42, 88 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Dance
PageRank 2 Idx 18 Total Score = 616.0 Four Factors = { 68, 42, 38, 78 } Company name = britannica URL https://www.britannica.com/art/dance
PageRank 3 Idx 16 Total Score = 521.0 Four Factors = { 44, 11, 73, 60 } Company name = dictionary URL https://www.dictionary.com/browse/dance
PageRank 4 Idx 0 Total Score = 489.0 Four Factors = { 47, 47, 30, 70 } Company name = bing URL https://www.bing.com/search?q=dance
PageRank 5 Idx 20 Total Score = 360.0 Four Factors = { 16, 39, 51, 49 } Company name = youtube URL https://www.youtube.com/watch?v=0pF9PN8UpqI
PageRank 6 Idx 11 Total Score = 357.0 Four Factors = { 0, 33, 61, 68 } Company name = youtube URL https://www.youtube.com/watch?v=MVhfaPoxBkw
PageRank 7 Idx 27 Total Score = 336.0 Four Factors = { 0, 34, 77, 37 } Company name = youtube URL https://www.youtube.com/watch?v=0u8ghlxK0c8
PageRank 8 Idx 9 Total Score = 327.0 Four Factors = { 5, 44, 42, 88 } Company name = theguardian URL https://www.theguardian.com/stage/dance
PageRank 9 Idx 4 Total Score = 319.0 Four Factors = { 0, 22, 71, 62 } Company name = youtube URL https://www.youtube.com/watch?v=2XtLEp0XAII
PageRank 10 Idx 6 Total Score = 312.0 Four Factors = { 1, 26, 63, 66 } Company name = sfgate URL https://www.sfgate.com/news/article/Axis-Dance-Ensemb
PageRank 11 Idx 29 Total Score = 309.0 Four Factors = { 0, 36, 47, 96 } Company name = youtube URL https://www.youtube.com/channel/UCi-g4ciqGV7ivU8aeS
PageRank 12 Idx 19 Total Score = 299.0 Four Factors = { 3, 38, 61, 25 } Company name = thoughtco URL https://www.thoughtco.com/dance-4132927
PageRank 13 Idx 8 Total Score = 299.0 Four Factors = { 0, 19, 52, 65 } Company name = instagram URL https://www.instagram.com/dance/
PageRank 14 Idx 26 Total Score = 298.0 Four Factors = { 0, 13, 66, 74 } Company name = google URL https://support.google.com/youtube/?hl=en
PageRank 15 Idx 3 Total Score = 297.0 Four Factors = { 1, 44, 53, 45 } Company name = youtube URL https://www.youtube.com/watch?v=2\_P0mghSrJ0
PageRank 16 Idx 14 Total Score = 289.0 Four Factors = { 0, 23, 54, 81 } Company name = youtube URL https://www.youtube.com/watch?v=iq010a5sWSU
PageRank 17 Idx 13 Total Score = 274.0 Four Factors = { 12, 37, 20, 80 } Company name = com/ URL https://dancestudio-pro.com/
PageRank 18 Idx 21 Total Score = 270.0 Four Factors = { 0, 39, 43, 63 } Company name = latimes URL http://www.latimes.com/entertainment/arts/theater/li
PageRank 19 Idx 5 Total Score = 257.0 Four Factors = { 0, 32, 37, 82 } Company name = youtube URL https://www.youtube.com/watch?v=qd6UI6wEIsU
PageRank 20 Idx 22 Total Score = 254.0 Four Factors = { 1, 26, 36, 89 } Company name = youtube URL https://www.youtube.com/watch?v=IczA3jvTx4c
PageRank 21 Idx 17 Total Score = 238.0 Four Factors = { 11, 7, 44, 37 } Company name = nytimes URL https://www.nytimes.com/section/arts/dance
PageRank 22 Idx 25 Total Score = 233.0 Four Factors = { 0, 47, 19, 82 } Company name = google URL https://plus.google.com/+youtube
PageRank 23 Idx 12 Total Score = 227.0 Four Factors = { 0, 12, 46, 15 } Company name = youtube URL https://www.youtube.com/watch?v=XYR0gJfXkc
PageRank 24 Idx 1 Total Score = 224.0 Four Factors = { 2, 12, 40, 70 } Company name = youtube URL https://www.youtube.com/watch?v=qd6hSLZkI7GY
PageRank 25 Idx 24 Total Score = 222.0 Four Factors = { 0, 44, 18, 80 } Company name = google URL https://www.google.com/intl/en/policies/privacy/
PageRank 26 Idx 7 Total Score = 214.0 Four Factors = { 4, 1, 44, 60 } Company name = popsugar URL https://www.popsugar.com/celebrity/Ted-Danson-Learn
PageRank 27 Idx 23 Total Score = 208.0 Four Factors = { 0, 21, 43, 37 } Company name = youtube URL https://www.youtube.com/yt/dev/
PageRank 28 Idx 15 Total Score = 201.0 Four Factors = { 0, 10, 50, 31 } Company name = youtube URL https://www.youtube.com/watch?v=Kl5B6MBaNTI
PageRank 29 Idx 28 Total Score = 185.0 Four Factors = { 0, 20, 19, 88 } Company name = youtube URL https://www.youtube.com/t/terms
PageRank 30 Idx 10 Total Score = 169.0 Four Factors = { 1, 6, 36, 44 } Company name = youtube URL https://www.youtube.com/watch?v=SoXGx7FakyU
```

INPUT: SUGAR

```
PageRank 1 Idx 13 Total Score = 923.0 Four Factors = { 138, 14, 67, 4 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Sugar
PageRank 2 Idx 7 Total Score = 489.0 Four Factors = { 69, 12, 27, 39 } Company name = merriam-webster URL https://www.merriam-webster.com/dictionary/sugar
PageRank 3 Idx 17 Total Score = 422.0 Four Factors = { 52, 48, 7, 45 } Company name = dictionary URL https://www.dictionary.com/browse/sugar
PageRank 4 Idx 9 Total Score = 395.0 Four Factors = { 0, 50, 68, 91 } Company name = youtube URL https://www.youtube.com/watch?v=DUX0D7w6lcs
PageRank 5 Idx 4 Total Score = 354.0 Four Factors = { 1, 0, 77, 100 } Company name = coolmathgames URL https://www.coolmathgames.com/0-sugar-sugar
PageRank 6 Idx 0 Total Score = 338.0 Four Factors = { 35, 16, 28, 47 } Company name = bing URL https://www.bing.com/search?q=sugar
PageRank 7 Idx 12 Total Score = 332.0 Four Factors = { 6, 28, 78, 12 } Company name = sugar URL https://www.sugar.org/
PageRank 8 Idx 24 Total Score = 328.0 Four Factors = { 0, 30, 72, 52 } Company name = webmd URL https://www.webmd.com/about-webmd-policies/about-privacy-poli
PageRank 9 Idx 3 Total Score = 313.0 Four Factors = { 0, 27, 63, 70 } Company name = net/record/421158 URL http://liverecorder.net/record/421158
PageRank 10 Idx 1 Total Score = 308.0 Four Factors = { 28, 20, 21, 65 } Company name = webmd URL https://www.webmd.com/diet/ss/slideshow-sugar-addiction
PageRank 11 Idx 21 Total Score = 290.0 Four Factors = { 0, 43, 57, 33 } Company name = webmd URL https://www.webmd.com/ed-treatment
PageRank 12 Idx 8 Total Score = 280.0 Four Factors = { 0, 24, 53, 73 } Company name = net/record/421222 URL http://liverecorder.net/record/421222
PageRank 13 Idx 19 Total Score = 278.0 Four Factors = { 2, 39, 42, 64 } Company name = webmd URL https://www.webmd.com/heart-disease/news/20181105/study-1-en
PageRank 14 Idx 18 Total Score = 276.0 Four Factors = { 0, 34, 59, 31 } Company name = webmd URL https://dictionary.webmd.com/default.htm
PageRank 15 Idx 25 Total Score = 269.0 Four Factors = { 0, 20, 57, 58 } Company name = webmd URL https://member.webmd.com/subscribe
PageRank 16 Idx 14 Total Score = 258.0 Four Factors = { 0, 18, 48, 78 } Company name = net/record/420682 URL http://liverecorder.net/record/420682
PageRank 17 Idx 28 Total Score = 256.0 Four Factors = { 0, 15, 71, 13 } Company name = webmd URL https://www.webmd.com/skin-problems-and-treatments/psoriasis
PageRank 18 Idx 22 Total Score = 243.0 Four Factors = { 0, 10, 66, 25 } Company name = webmd URL https://www.webmd.com/skin-problems-and-treatments/default.h
PageRank 19 Idx 5 Total Score = 210.0 Four Factors = { 0, 24, 27, 81 } Company name = youtube URL https://www.youtube.com/watch?v=5vBG0r16y8k
PageRank 20 Idx 11 Total Score = 193.0 Four Factors = { 0, 24, 43, 16 } Company name = net/record/421611 URL http://liverecorder.net/record/421611
PageRank 21 Idx 16 Total Score = 183.0 Four Factors = { 0, 24, 37, 24 } Company name = net/record/421694 URL https://liverecorder.net/record/421694
PageRank 22 Idx 2 Total Score = 175.0 Four Factors = { 0, 6, 37, 52 } Company name = net/record/420672 URL http://liverecorder.net/record/420672
PageRank 23 Idx 15 Total Score = 174.0 Four Factors = { 4, 27, 24, 28 } Company name = abcyu URL http://www.abcyu.com/sugar\_sugar\_2.htm
PageRank 24 Idx 27 Total Score = 169.0 Four Factors = { 0, 22, 9, 98 } Company name = webmd URL https://www.webmd.com/special-reports/opioids-pain/default.ht
PageRank 25 Idx 20 Total Score = 169.0 Four Factors = { 0, 6, 46, 19 } Company name = webmd URL https://www.webmd.com/mental-health/news/20181114/yes-another
PageRank 26 Idx 10 Total Score = 161.0 Four Factors = { 1, 19, 10, 88 } Company name = youtube URL https://www.youtube.com/watch?v=9nE2sp0w\_o
PageRank 27 Idx 6 Total Score = 127.0 Four Factors = { 0, 33, 12, 25 } Company name = youtube URL https://www.youtube.com/watch?v=9nE2sp0w\_o
PageRank 28 Idx 23 Total Score = 121.0 Four Factors = { 0, 9, 21, 40 } Company name = webmd URL https://www.webmd.com/fitness-exercise/ss/slideshow-exercise-
PageRank 29 Idx 26 Total Score = 88.0 Four Factors = { 0, 7, 20, 14 } Company name = webmd URL https://www.webmd.com/heart-disease/default.htm
PageRank 30 Idx 29 Total Score = 86.0 Four Factors = { 0, 12, 12, 26 } Company name = webmd URL https://www.webmd.com/parenting/diapering-skin-care-17/video-
```

INPUT: CAT

PageRank 1 Idx 21 Total Score = 1111.0 Four Factors = { 178, 38, 32, 49 } Company name = wikipedia URL <https://en.wikipedia.org/wiki/Cat>

PageRank 2 Idx 7 Total Score = 504.0 Four Factors = { 72, 41, 4, 50 } Company name = dictionary URL <https://www.dictionary.com/browse/cat>

PageRank 3 Idx 29 Total Score = 337.0 Four Factors = { 0, 43, 55, 86 } Company name = cat URL [https://www.cat.com/tr\\_TR.html](https://www.cat.com/tr_TR.html)

PageRank 4 Idx 13 Total Score = 313.0 Four Factors = { 3, 14, 69, 63 } Company name = youtube URL <https://www.youtube.com/watch?v=1Wh8RzcQZr4>

PageRank 5 Idx 24 Total Score = 301.0 Four Factors = { 0, 32, 77, 6 } Company name = cat URL [https://www.cat.com/ko\\_KR.html](https://www.cat.com/ko_KR.html)

PageRank 6 Idx 28 Total Score = 300.0 Four Factors = { 0, 43, 71, 1 } Company name = cat URL [https://www.cat.com/en\\_MX.html](https://www.cat.com/en_MX.html)

PageRank 7 Idx 18 Total Score = 290.0 Four Factors = { 0, 3, 72, 68 } Company name = cat URL [https://www.cat.com/en\\_US/campaigns/awareness/cat-fluids.h](https://www.cat.com/en_US/campaigns/awareness/cat-fluids.h)

PageRank 8 Idx 0 Total Score = 289.0 Four Factors = { 34, 24, 7, 50 } Company name = bing URL <https://www.bing.com/search?q=cat>

PageRank 9 Idx 23 Total Score = 281.0 Four Factors = { 0, 39, 39, 86 } Company name = cat URL [https://www.cat.com/fr\\_US.html](https://www.cat.com/fr_US.html)

PageRank 10 Idx 20 Total Score = 274.0 Four Factors = { 0, 6, 59, 85 } Company name = cat URL [https://www.cat.com/en\\_US/campaigns/awareness/excavator-a](https://www.cat.com/en_US/campaigns/awareness/excavator-a)

PageRank 11 Idx 1 Total Score = 271.0 Four Factors = { 0, 25, 46, 83 } Company name = cat URL <https://www.cat.com/>

PageRank 12 Idx 12 Total Score = 266.0 Four Factors = { 0, 31, 59, 27 } Company name = cat URL <https://parts.cat.com/en/catcorp/>

PageRank 13 Idx 16 Total Score = 253.0 Four Factors = { 0, 45, 27, 82 } Company name = marketwatch URL <https://www.marketwatch.com/tools/markets/stocks>

PageRank 14 Idx 10 Total Score = 237.0 Four Factors = { 0, 18, 50, 51 } Company name = marketwatch URL <https://www.marketwatch.com/tools/markets>

PageRank 15 Idx 6 Total Score = 235.0 Four Factors = { 0, 48, 38, 25 } Company name = cat URL [https://www.cat.com/en\\_US/campaigns/awareness/attachments](https://www.cat.com/en_US/campaigns/awareness/attachments)

PageRank 16 Idx 8 Total Score = 232.0 Four Factors = { 0, 32, 23, 99 } Company name = yahoo URL <https://finance.yahoo.com/quote/CAT>

PageRank 17 Idx 26 Total Score = 217.0 Four Factors = { 0, 35, 25, 72 } Company name = cat URL [https://www.cat.com/en\\_ZA.html](https://www.cat.com/en_ZA.html)

PageRank 18 Idx 15 Total Score = 217.0 Four Factors = { 0, 26, 36, 57 } Company name = cat URL [https://www.cat.com/en\\_US/campaigns/awareness/hammers-at](https://www.cat.com/en_US/campaigns/awareness/hammers-at)

PageRank 19 Idx 3 Total Score = 214.0 Four Factors = { 7, 39, 28, 17 } Company name = vetstreet URL <http://www.vetstreet.com/cats/>

PageRank 20 Idx 22 Total Score = 189.0 Four Factors = { 0, 40, 18, 55 } Company name = marketwatch URL <https://www.marketwatch.com/tools/markets/stocks>

PageRank 21 Idx 2 Total Score = 174.0 Four Factors = { 0, 43, 23, 19 } Company name = cnn URL <https://money.cnn.com/quote/quote.html?symb=CAT>

PageRank 22 Idx 14 Total Score = 171.0 Four Factors = { 4, 6, 41, 16 } Company name = youtube URL <https://www.youtube.com/watch?v=opYts0nF0vs>

PageRank 23 Idx 11 Total Score = 166.0 Four Factors = { 7, 4, 15, 78 } Company name = charlotteobserver URL <https://www.charlotteobserver.com/news/stat>

PageRank 24 Idx 9 Total Score = 165.0 Four Factors = { 1, 37, 8, 62 } Company name = youtube URL <https://www.youtube.com/watch?v=yWlRvhfipI8>

PageRank 25 Idx 25 Total Score = 157.0 Four Factors = { 0, 12, 35, 28 } Company name = googletagmanager URL <https://www.googletagmanager.com/ns.html?id>

PageRank 26 Idx 19 Total Score = 137.0 Four Factors = { 0, 19, 23, 30 } Company name = cat URL [https://www.cat.com/en\\_US/campaigns/awareness/myaccount.i](https://www.cat.com/en_US/campaigns/awareness/myaccount.i)

PageRank 27 Idx 17 Total Score = 113.0 Four Factors = { 3, 35, 1, 25 } Company name = youtube URL <https://www.youtube.com/watch?v=3iyMaW0ZGoA>

PageRank 28 Idx 5 Total Score = 96.0 Four Factors = { 0, 1, 18, 40 } Company name = cat URL [https://www.cat.com/en\\_US/campaigns/awareness/shears-attach](https://www.cat.com/en_US/campaigns/awareness/shears-attach)

PageRank 29 Idx 27 Total Score = 78.0 Four Factors = { 0, 2, 8, 50 } Company name = cat URL [https://www.cat.com/ru\\_RU.html](https://www.cat.com/ru_RU.html)

PageRank 30 Idx 4 Total Score = 57.0 Four Factors = { 0, 21, 4, 3 } Company name = cat URL [https://www.cat.com/en\\_US/by-industry/mining/OperatorTrainin](https://www.cat.com/en_US/by-industry/mining/OperatorTrainin)

## INPUT: TEA

PageRank 1 Idx 7 Total Score = 2010.0 Four Factors = { 370, 41, 11, 45 } Company name = wikipedia URL <https://en.wikipedia.org/wiki/Tea>

PageRank 2 Idx 0 Total Score = 472.0 Four Factors = { 50, 32, 35, 53 } Company name = bing URL <https://www.bing.com/search?q=tea>

PageRank 3 Idx 4 Total Score = 420.0 Four Factors = { 20, 34, 75, 27 } Company name = jazentea URL <http://www.jazentea.com/>

PageRank 4 Idx 3 Total Score = 418.0 Four Factors = { 40, 35, 49, 1 } Company name = yelp URL [https://www.yelp.com/biz/k-tea-cafe-sunnyvale?adjust\\_creat](https://www.yelp.com/biz/k-tea-cafe-sunnyvale?adjust_creat)

PageRank 5 Idx 17 Total Score = 377.0 Four Factors = { 56, 9, 11, 46 } Company name = yelp URL [https://www.yelp.com/biz/jazen-tea-santa-clara-3?adjust\\_creat](https://www.yelp.com/biz/jazen-tea-santa-clara-3?adjust_creat)

PageRank 6 Idx 8 Total Score = 377.0 Four Factors = { 0, 31, 80, 75 } Company name = teacollection URL <https://www.teacollection.com/>

PageRank 7 Idx 11 Total Score = 369.0 Four Factors = { 0, 44, 70, 71 } Company name = com/sunnyvalewest URL <http://quicklyusa.com/sunnyvalewest.html>

PageRank 8 Idx 12 Total Score = 350.0 Four Factors = { 58, 18, 3, 15 } Company name = britannica URL <https://www.britannica.com/topic/tea-beverage>

PageRank 9 Idx 14 Total Score = 347.0 Four Factors = { 0, 36, 67, 74 } Company name = texas URL <https://tea.texas.gov/student/assessment/staar/>

PageRank 10 Idx 19 Total Score = 341.0 Four Factors = { 36, 3, 28, 71 } Company name = yelp URL [https://www.yelp.com/biz/quickly-sunnyvale-4?adjust\\_creat](https://www.yelp.com/biz/quickly-sunnyvale-4?adjust_creat)

PageRank 11 Idx 9 Total Score = 331.0 Four Factors = { 0, 14, 77, 72 } Company name = texas URL <https://tea.texas.gov/About-TEA/Contact/>

PageRank 12 Idx 23 Total Score = 328.0 Four Factors = { 0, 47, 75, 9 } Company name = govdelivery URL <https://public.govdelivery.com/accounts/TXTEA/subs>

PageRank 13 Idx 29 Total Score = 323.0 Four Factors = { 0, 27, 68, 65 } Company name = tea URL <https://tealprod.tea.state.tx.us/fsp/Reports/ReportSelect>

PageRank 14 Idx 21 Total Score = 317.0 Four Factors = { 0, 49, 45, 84 } Company name = tea URL <http://ritter.tea.state.tx.us/rules/tac/chapter126/index>

PageRank 15 Idx 6 Total Score = 288.0 Four Factors = { 0, 33, 56, 54 } Company name = stashtea URL <https://www.stashtea.com/>

PageRank 16 Idx 27 Total Score = 273.0 Four Factors = { 0, 21, 50, 81 } Company name = tea URL <http://ritter.tea.state.tx.us/rules/tac/chapter111/index>

PageRank 17 Idx 18 Total Score = 272.0 Four Factors = { 25, 13, 26, 43 } Company name = webmd URL <https://www.webmd.com/diet/features/tea-types-and-thei>

PageRank 18 Idx 2 Total Score = 234.0 Four Factors = { 4, 45, 11, 91 } Company name = republicoftea URL <https://www.republicoftea.com/>

PageRank 19 Idx 13 Total Score = 232.0 Four Factors = { 0, 23, 40, 66 } Company name = texas URL <https://tea.texas.gov/Texas-Educators/>

PageRank 20 Idx 10 Total Score = 231.0 Four Factors = { 0, 10, 68, 7 } Company name = com/teainfo URL <https://twitter.com/teainfo>

PageRank 21 Idx 20 Total Score = 219.0 Four Factors = { 0, 41, 40, 17 } Company name = yelp URL [https://www.yelp.com/biz/starbucks-sunnyvale?adjust\\_creat](https://www.yelp.com/biz/starbucks-sunnyvale?adjust_creat)

PageRank 22 Idx 1 Total Score = 209.0 Four Factors = { 0, 34, 21, 78 } Company name = texas URL <https://tea.texas.gov/curriculum/teks/>

PageRank 23 Idx 22 Total Score = 208.0 Four Factors = { 0, 44, 21, 57 } Company name = tea URL <http://ritter.tea.state.tx.us/rules/tac/chapter116/index>

PageRank 24 Idx 25 Total Score = 193.0 Four Factors = { 0, 49, 2, 89 } Company name = tea URL <https://tealprod.tea.state.tx.us/WaiversReports/Tea.Waiver>

PageRank 25 Idx 28 Total Score = 182.0 Four Factors = { 0, 38, 21, 43 } Company name = texasgateway URL <https://www.texasgateway.org/>

PageRank 26 Idx 5 Total Score = 168.0 Four Factors = { 2, 43, 12, 36 } Company name = yelp URL <https://www.yelp.com/biz/peets-coffee-and-tea-sunnyvale-3>

PageRank 27 Idx 16 Total Score = 167.0 Four Factors = { 0, 2, 33, 64 } Company name = texas URL <https://tea.texas.gov/Texas-Educators/Certification/Cert>

PageRank 28 Idx 15 Total Score = 135.0 Four Factors = { 0, 9, 6, 99 } Company name = texas URL <https://tea.texas.gov/Home/>

PageRank 29 Idx 26 Total Score = 113.0 Four Factors = { 0, 24, 15, 20 } Company name = texas URL <https://tea.texas.gov/sitemap/>

PageRank 30 Idx 24 Total Score = 56.0 Four Factors = { 0, 4, 14, 6 } Company name = tea URL [http://tea4avholly.tea.state.tx.us/tea\\_asktd.web/Forms/Home](http://tea4avholly.tea.state.tx.us/tea_asktd.web/Forms/Home)

## INPUT: ROCK



PageRank 1 Idx 19 Total Score = 2132.0 Four Factors = { 364, 47, 69, 11 } Company name = thefreedictionary URL <https://www.thefreedictionary.com/rock>  
PageRank 2 Idx 15 Total Score = 512.0 Four Factors = { 69, 47, 14, 31 } Company name = merriam-webster URL <https://www.merriam-webster.com/dictionary>  
PageRank 3 Idx 5 Total Score = 384.0 Four Factors = { 28, 16, 51, 59 } Company name = wikipedia URL [https://en.wikipedia.org/wiki/Rock\\_\(geology\)](https://en.wikipedia.org/wiki/Rock_(geology))  
PageRank 4 Idx 8 Total Score = 328.0 Four Factors = { 2, 23, 59, 95 } Company name = youtube URL <https://www.youtube.com/watch?v=htlMqRfL5A>  
PageRank 5 Idx 7 Total Score = 310.0 Four Factors = { 1, 18, 63, 80 } Company name = bbc URL <https://www.bbc.co.uk/news/uk-england-nottinghamshire-46>  
PageRank 6 Idx 0 Total Score = 310.0 Four Factors = { 16, 8, 56, 46 } Company name = bing URL <https://www.bing.com/search?q=rock>  
PageRank 7 Idx 10 Total Score = 308.0 Four Factors = { 0, 20, 57, 97 } Company name = foxnews URL <https://www.foxnews.com/entertainment/jack-black-re>  
PageRank 8 Idx 25 Total Score = 306.0 Four Factors = { 0, 43, 71, 7 } Company name = rockauto URL <https://www.rockauto.com/en/tools/?parttype=260>  
PageRank 9 Idx 20 Total Score = 300.0 Four Factors = { 0, 14, 58, 98 } Company name = bing URL <http://help.bing.microsoft.com/>  
PageRank 10 Idx 23 Total Score = 292.0 Four Factors = { 0, 50, 38, 78 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2013>  
PageRank 11 Idx 27 Total Score = 277.0 Four Factors = { 0, 24, 71, 16 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,1963>  
PageRank 12 Idx 13 Total Score = 269.0 Four Factors = { 0, 19, 68, 27 } Company name = youtube URL <https://www.youtube.com/watch?v=9Xa593WMRQ0>  
PageRank 13 Idx 18 Total Score = 262.0 Four Factors = { 0, 28, 66, 8 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/dodge>  
PageRank 14 Idx 6 Total Score = 255.0 Four Factors = { 14, 36, 36, 5 } Company name = yelp URL <https://www.yelp.com/biz/rok-bistro-sunnyvale-2>  
PageRank 15 Idx 21 Total Score = 253.0 Four Factors = { 0, 15, 68, 19 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2011>  
PageRank 16 Idx 1 Total Score = 250.0 Four Factors = { 0, 13, 57, 53 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford>  
PageRank 17 Idx 22 Total Score = 250.0 Four Factors = { 0, 1, 52, 92 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2010>  
PageRank 18 Idx 26 Total Score = 249.0 Four Factors = { 0, 30, 46, 51 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2019>  
PageRank 19 Idx 4 Total Score = 249.0 Four Factors = { 0, 47, 25, 80 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/toyota>  
PageRank 20 Idx 14 Total Score = 222.0 Four Factors = { 0, 10, 52, 46 } Company name = rockauto URL <https://www.rockauto.com/>  
PageRank 21 Idx 3 Total Score = 217.0 Four Factors = { 0, 2, 50, 63 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/>  
PageRank 22 Idx 16 Total Score = 205.0 Four Factors = { 0, 16, 31, 80 } Company name = youtube URL <https://www.youtube.com/watch?v=85bUC-UaAxE>  
PageRank 23 Idx 29 Total Score = 190.0 Four Factors = { 0, 40, 28, 26 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,1964>  
PageRank 24 Idx 12 Total Score = 188.0 Four Factors = { 2, 20, 15, 93 } Company name = latimes URL <http://www.latimes.com/books/la-ca-ic-jeff-tweedy->  
PageRank 25 Idx 28 Total Score = 175.0 Four Factors = { 0, 11, 37, 42 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2018>  
PageRank 26 Idx 2 Total Score = 167.0 Four Factors = { 0, 7, 18, 99 } Company name = youtube URL <https://www.youtube.com/watch?v=tVv6SrRI9pU>  
PageRank 27 Idx 11 Total Score = 166.0 Four Factors = { 0, 8, 34, 48 } Company name = com/ URL <http://rockchurchsunnyvale.com/>  
PageRank 28 Idx 9 Total Score = 154.0 Four Factors = { 0, 22, 23, 41 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/jeep>  
PageRank 29 Idx 17 Total Score = 77.0 Four Factors = { 0, 14, 8, 25 } Company name = youtube URL <https://www.youtube.com/watch?v=iw70gIMMRc4>  
PageRank 30 Idx 24 Total Score = 34.0 Four Factors = { 0, 7, 5, 5 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2012>

## TEST BST SEARCH:

### SEARCH RANK 3:

PageRank 1 Idx 19 Total Score = 2132.0 Four Factors = { 364, 47, 69, 11 } Company name = thefreedictionary URL <https://www.thefreedictionary.com>  
PageRank 2 Idx 15 Total Score = 512.0 Four Factors = { 69, 47, 14, 31 } Company name = merriam-webster URL <https://www.merriam-webster.com/dicti>  
PageRank 3 Idx 5 Total Score = 384.0 Four Factors = { 28, 16, 51, 59 } Company name = wikipedia URL [https://en.wikipedia.org/wiki/Rock\\_\(geology\)](https://en.wikipedia.org/wiki/Rock_(geology))  
PageRank 4 Idx 8 Total Score = 328.0 Four Factors = { 2, 23, 59, 95 } Company name = youtube URL <https://www.youtube.com/watch?v=htlMqRfL5A>  
PageRank 5 Idx 7 Total Score = 310.0 Four Factors = { 1, 18, 63, 80 } Company name = bbc URL <https://www.bbc.co.uk/news/uk-england-nottinghamshi>  
PageRank 6 Idx 0 Total Score = 310.0 Four Factors = { 16, 8, 56, 46 } Company name = bing URL <https://www.bing.com/search?q=rock>  
PageRank 7 Idx 10 Total Score = 308.0 Four Factors = { 0, 20, 57, 97 } Company name = foxnews URL <https://www.foxnews.com/entertainment/jack-bla>  
PageRank 8 Idx 25 Total Score = 306.0 Four Factors = { 0, 43, 71, 7 } Company name = rockauto URL <https://www.rockauto.com/en/tools/?parttype=260>  
PageRank 9 Idx 20 Total Score = 300.0 Four Factors = { 0, 14, 58, 98 } Company name = bing URL <http://help.bing.microsoft.com/>  
PageRank 10 Idx 23 Total Score = 292.0 Four Factors = { 0, 50, 38, 78 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2013>  
PageRank 11 Idx 27 Total Score = 277.0 Four Factors = { 0, 24, 71, 16 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,1963>  
PageRank 12 Idx 13 Total Score = 269.0 Four Factors = { 0, 19, 68, 27 } Company name = youtube URL <https://www.youtube.com/watch?v=9Xa593WMRQ0>  
PageRank 13 Idx 18 Total Score = 262.0 Four Factors = { 0, 28, 66, 8 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/dodge>  
PageRank 14 Idx 6 Total Score = 255.0 Four Factors = { 14, 36, 36, 5 } Company name = yelp URL <https://www.yelp.com/biz/rok-bistro-sunnyvale-2>  
PageRank 15 Idx 21 Total Score = 253.0 Four Factors = { 0, 15, 68, 19 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2011>  
PageRank 16 Idx 1 Total Score = 250.0 Four Factors = { 0, 13, 57, 53 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford>  
PageRank 17 Idx 22 Total Score = 250.0 Four Factors = { 0, 1, 52, 92 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2010>  
PageRank 18 Idx 26 Total Score = 249.0 Four Factors = { 0, 30, 46, 51 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2019>  
PageRank 19 Idx 4 Total Score = 249.0 Four Factors = { 0, 47, 25, 80 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/toyota>  
PageRank 20 Idx 14 Total Score = 222.0 Four Factors = { 0, 10, 52, 46 } Company name = rockauto URL <https://www.rockauto.com/>  
PageRank 21 Idx 3 Total Score = 217.0 Four Factors = { 0, 2, 50, 63 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/>  
PageRank 22 Idx 16 Total Score = 205.0 Four Factors = { 0, 16, 31, 80 } Company name = youtube URL <https://www.youtube.com/watch?v=85bUC-UaAxE>  
PageRank 23 Idx 29 Total Score = 190.0 Four Factors = { 0, 40, 28, 26 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,1964>  
PageRank 24 Idx 12 Total Score = 188.0 Four Factors = { 2, 20, 15, 93 } Company name = latimes URL <http://www.latimes.com/books/la-ca-ic-jeff-tw>  
PageRank 25 Idx 28 Total Score = 175.0 Four Factors = { 0, 11, 37, 42 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2018>  
PageRank 26 Idx 2 Total Score = 167.0 Four Factors = { 0, 7, 18, 99 } Company name = youtube URL <https://www.youtube.com/watch?v=tVv6SrRI9pU>  
PageRank 27 Idx 11 Total Score = 166.0 Four Factors = { 0, 8, 34, 48 } Company name = com/ URL <http://rockchurchsunnyvale.com/>  
PageRank 28 Idx 9 Total Score = 154.0 Four Factors = { 0, 22, 23, 41 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/jeep>  
PageRank 29 Idx 17 Total Score = 77.0 Four Factors = { 0, 14, 8, 25 } Company name = youtube URL <https://www.youtube.com/watch?v=iw70gIMMRc4>  
PageRank 30 Idx 24 Total Score = 34.0 Four Factors = { 0, 7, 5, 5 } Company name = rockauto URL <https://www.rockauto.com/en/catalog/ford,2012>  
Test Process BST? Y/N: Y

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;  
1  
Enter a Page Rank:  
3  
3 Idx 5 Total Score = 384.0 Four Factors = { 28, 16, 51, 59 } Company name = wikipedia URL [https://en.wikipedia.org/wiki/Rock\\_\(geology\)](https://en.wikipedia.org/wiki/Rock_(geology))

### SEARCH RANK 5:

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;  
1  
Enter a Page Rank:  
5  
5 Idx 7 Total Score = 310.0 Four Factors = { 1, 18, 63, 80 } Company name = bbc URL <https://www.bbc.co.uk/news/uk-england-nottinghamshire-46244014>

### SEARCH RANK 20:

```
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
1
Enter a Page Rank:
20
20 Idx 14 Total Score = 222.0 Four Factors = { 0, 10, 52, 46 } Company name = rockauto URL https://www.rockauto.com/
```

## TEST BST DELETION:

### DELETE RANK 5

URL: <https://www.bbc.co.uk/news/uk-england-nottinghamshire-46244014> :

```
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
3
Enter a Page Rank:
5

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
4
1 Idx 19 Total Score = 2132.0 Four Factors = { 364, 47, 69, 11 } Company name = thefreedictionary URL https://www.thefreedictionary.com/rock
2 Idx 15 Total Score = 512.0 Four Factors = { 69, 47, 14, 31 } Company name = merriam-webster URL https://www.merriam-webster.com/dictionary/rock
3 Idx 5 Total Score = 384.0 Four Factors = { 28, 16, 51, 59 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Rock\_\(geology\)
4 Idx 8 Total Score = 328.0 Four Factors = { 2, 23, 59, 95 } Company name = youtube URL https://www.youtube.com/watch?v=htlMqRfL5A
6 Idx 0 Total Score = 310.0 Four Factors = { 16, 8, 56, 46 } Company name = bing URL https://www.bing.com/search?q=rock
7 Idx 10 Total Score = 308.0 Four Factors = { 0, 20, 57, 97 } Company name = foxnews URL https://www.foxnews.com/entertainment/jack-black-reunites-
8 Idx 25 Total Score = 306.0 Four Factors = { 0, 43, 71, 7 } Company name = rockauto URL https://www.rockauto.com/en/tools/?parttype=260
9 Idx 20 Total Score = 300.0 Four Factors = { 0, 14, 58, 98 } Company name = bing URL http://help.bing.microsoft.com/
10 Idx 23 Total Score = 292.0 Four Factors = { 0, 50, 38, 78 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2013
11 Idx 27 Total Score = 277.0 Four Factors = { 0, 24, 71, 16 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1963
12 Idx 13 Total Score = 269.0 Four Factors = { 0, 19, 68, 27 } Company name = youtube URL https://www.youtube.com/watch?v=9Xa593wMR0Q
13 Idx 18 Total Score = 262.0 Four Factors = { 0, 28, 66, 8 } Company name = rockauto URL https://www.rockauto.com/en/catalog/dodge
14 Idx 6 Total Score = 255.0 Four Factors = { 14, 36, 36, 5 } Company name = yelp URL https://www.yelp.com/biz/rok-bistro-sunnyvale-2
15 Idx 21 Total Score = 253.0 Four Factors = { 0, 15, 68, 19 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2011
16 Idx 1 Total Score = 250.0 Four Factors = { 0, 13, 57, 53 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford
17 Idx 22 Total Score = 250.0 Four Factors = { 0, 1, 52, 92 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2010
18 Idx 26 Total Score = 249.0 Four Factors = { 0, 30, 46, 51 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2019
19 Idx 4 Total Score = 249.0 Four Factors = { 0, 47, 25, 80 } Company name = rockauto URL https://www.rockauto.com/en/catalog/toyota
20 Idx 14 Total Score = 222.0 Four Factors = { 0, 10, 52, 46 } Company name = rockauto URL https://www.rockauto.com/
21 Idx 3 Total Score = 217.0 Four Factors = { 0, 2, 50, 63 } Company name = rockauto URL https://www.rockauto.com/en/catalog/
22 Idx 16 Total Score = 205.0 Four Factors = { 0, 16, 31, 80 } Company name = youtube URL https://www.youtube.com/watch?v=8SbUC-UaAxE
23 Idx 29 Total Score = 190.0 Four Factors = { 0, 40, 28, 26 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1964
24 Idx 12 Total Score = 188.0 Four Factors = { 2, 20, 15, 93 } Company name = latimes URL http://www.latimes.com/books/la-ca-ic-jeff-tweedy-20181116
25 Idx 28 Total Score = 175.0 Four Factors = { 0, 11, 37, 42 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2018
26 Idx 2 Total Score = 167.0 Four Factors = { 0, 7, 18, 99 } Company name = youtube URL https://www.youtube.com/watch?v=tVv6SrRI9pU
27 Idx 11 Total Score = 166.0 Four Factors = { 0, 8, 34, 48 } Company name = com/ URL http://rockchurchsunnyvale.com/
28 Idx 9 Total Score = 154.0 Four Factors = { 0, 22, 23, 41 } Company name = rockauto URL https://www.rockauto.com/en/catalog/jeep
29 Idx 17 Total Score = 77.0 Four Factors = { 0, 14, 8, 25 } Company name = youtube URL https://www.youtube.com/watch?v=iw70gIMMRc4
30 Idx 24 Total Score = 34.0 Four Factors = { 0, 7, 5, 5 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2012

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
```

### DELETE RANK 8

```
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
3
Enter a Page Rank:
8

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
4
1 Idx 19 Total Score = 2132.0 Four Factors = { 364, 47, 69, 11 } Company name = thefreedictionary URL https://www.thefreedictionary.com/rock
2 Idx 15 Total Score = 512.0 Four Factors = { 69, 47, 14, 31 } Company name = merriam-webster URL https://www.merriam-webster.com/dictionary/rock
3 Idx 5 Total Score = 384.0 Four Factors = { 28, 16, 51, 59 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Rock\_\(geology\)
4 Idx 8 Total Score = 328.0 Four Factors = { 2, 23, 59, 95 } Company name = youtube URL https://www.youtube.com/watch?v=htlMqRfL5A
6 Idx 0 Total Score = 310.0 Four Factors = { 16, 8, 56, 46 } Company name = bing URL https://www.bing.com/search?q=rock
7 Idx 10 Total Score = 308.0 Four Factors = { 0, 20, 57, 97 } Company name = foxnews URL https://www.foxnews.com/entertainment/jack-black-reunites-
9 Idx 20 Total Score = 300.0 Four Factors = { 0, 14, 58, 98 } Company name = bing URL http://help.bing.microsoft.com/
10 Idx 23 Total Score = 292.0 Four Factors = { 0, 50, 38, 78 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2013
11 Idx 27 Total Score = 277.0 Four Factors = { 0, 24, 71, 16 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1963
12 Idx 13 Total Score = 269.0 Four Factors = { 0, 19, 68, 27 } Company name = youtube URL https://www.youtube.com/watch?v=9Xa593wMR0Q
13 Idx 18 Total Score = 262.0 Four Factors = { 0, 28, 66, 8 } Company name = rockauto URL https://www.rockauto.com/en/catalog/dodge
14 Idx 6 Total Score = 255.0 Four Factors = { 14, 36, 36, 5 } Company name = yelp URL https://www.yelp.com/biz/rok-bistro-sunnyvale-2
15 Idx 21 Total Score = 253.0 Four Factors = { 0, 15, 68, 19 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2011
16 Idx 1 Total Score = 250.0 Four Factors = { 0, 13, 57, 53 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford
17 Idx 22 Total Score = 250.0 Four Factors = { 0, 1, 52, 92 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2010
18 Idx 26 Total Score = 249.0 Four Factors = { 0, 30, 46, 51 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2019
19 Idx 4 Total Score = 249.0 Four Factors = { 0, 47, 25, 80 } Company name = rockauto URL https://www.rockauto.com/en/catalog/toyota
20 Idx 14 Total Score = 222.0 Four Factors = { 0, 10, 52, 46 } Company name = rockauto URL https://www.rockauto.com/
21 Idx 3 Total Score = 217.0 Four Factors = { 0, 2, 50, 63 } Company name = rockauto URL https://www.rockauto.com/en/catalog/
22 Idx 16 Total Score = 205.0 Four Factors = { 0, 16, 31, 80 } Company name = youtube URL https://www.youtube.com/watch?v=8SbUC-UaAxE
23 Idx 29 Total Score = 190.0 Four Factors = { 0, 40, 28, 26 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1964
24 Idx 12 Total Score = 188.0 Four Factors = { 2, 20, 15, 93 } Company name = latimes URL http://www.latimes.com/books/la-ca-ic-jeff-tweedy-20181116
25 Idx 28 Total Score = 175.0 Four Factors = { 0, 11, 37, 42 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2018
26 Idx 2 Total Score = 167.0 Four Factors = { 0, 7, 18, 99 } Company name = youtube URL https://www.youtube.com/watch?v=tVv6SrRI9pU
27 Idx 11 Total Score = 166.0 Four Factors = { 0, 8, 34, 48 } Company name = com/ URL http://rockchurchsunnyvale.com/
28 Idx 9 Total Score = 154.0 Four Factors = { 0, 22, 23, 41 } Company name = rockauto URL https://www.rockauto.com/en/catalog/jeep
29 Idx 17 Total Score = 77.0 Four Factors = { 0, 14, 8, 25 } Company name = youtube URL https://www.youtube.com/watch?v=iw70gIMMRc4
30 Idx 24 Total Score = 34.0 Four Factors = { 0, 7, 5, 5 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2012
```



## TEST BST INSERTION:

### Insert rank 2

URL: [www.cat.com](http://www.cat.com)

```
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
2
Enter a Page Rank:
6
Enter a URL:
www.cat.com

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
4
1 Idx 19 Total Score = 2132.0 Four Factors = { 364, 47, 69, 11 } Company name = thefreedictionary URL https://www.thefreedictionary.com/rock
2 Idx 15 Total Score = 512.0 Four Factors = { 69, 47, 14, 31 } Company name = merriam-webster URL https://www.merriam-webster.com/dictionary/rock
3 Idx 5 Total Score = 384.0 Four Factors = { 28, 16, 51, 59 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Rock\_\(geology\)
4 Idx 8 Total Score = 328.0 Four Factors = { 2, 23, 59, 95 } Company name = youtube URL https://www.youtube.com/watch?v=htlnQrFL5A
6 Idx 0 Total Score = 310.0 Four Factors = { 16, 8, 56, 46 } Company name = bing URL https://www.bing.com/search?q=rock
7 Idx 0 Total Score = 0.0 Four Factors = { 0, 0, 0, 0 } Company name = cat URL www.cat.com
7 Idx 10 Total Score = 308.0 Four Factors = { 0, 20, 57, 97 } Company name = foxnews URL https://www.foxnews.com/entertainment/jack-black-reunites
9 Idx 20 Total Score = 300.0 Four Factors = { 0, 14, 58, 98 } Company name = bing URL http://help.bing.microsoft.com/
10 Idx 23 Total Score = 292.0 Four Factors = { 0, 50, 38, 78 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2013
11 Idx 27 Total Score = 277.0 Four Factors = { 0, 24, 71, 16 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1963
12 Idx 13 Total Score = 269.0 Four Factors = { 0, 19, 68, 27 } Company name = youtube URL https://www.youtube.com/watch?v=9XaS93wMR0Q
13 Idx 18 Total Score = 262.0 Four Factors = { 0, 28, 66, 8 } Company name = rockauto URL https://www.rockauto.com/en/catalog/dodge
14 Idx 6 Total Score = 255.0 Four Factors = { 14, 36, 36, 5 } Company name = yelp URL https://www.yelp.com/biz/rok-bistro-sunnysvale-2
15 Idx 21 Total Score = 253.0 Four Factors = { 0, 15, 68, 19 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2011
16 Idx 1 Total Score = 250.0 Four Factors = { 0, 13, 57, 53 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford
17 Idx 22 Total Score = 250.0 Four Factors = { 0, 1, 52, 92 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2010
18 Idx 26 Total Score = 249.0 Four Factors = { 0, 30, 46, 51 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2019
19 Idx 4 Total Score = 249.0 Four Factors = { 0, 47, 25, 80 } Company name = rockauto URL https://www.rockauto.com/en/catalog/toyota
20 Idx 14 Total Score = 222.0 Four Factors = { 0, 10, 52, 46 } Company name = rockauto URL https://www.rockauto.com/
21 Idx 3 Total Score = 217.0 Four Factors = { 0, 2, 50, 63 } Company name = rockauto URL https://www.rockauto.com/en/catalog/
22 Idx 16 Total Score = 205.0 Four Factors = { 0, 16, 31, 80 } Company name = youtube URL https://www.youtube.com/watch?v=8SbUC-UaXfE
23 Idx 29 Total Score = 190.0 Four Factors = { 0, 40, 28, 26 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1964
24 Idx 12 Total Score = 188.0 Four Factors = { 2, 20, 15, 93 } Company name = latimes URL http://www.latimes.com/books/la-ca-ic-jeff-tweedy-201811
25 Idx 28 Total Score = 175.0 Four Factors = { 0, 11, 37, 42 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2018
26 Idx 2 Total Score = 167.0 Four Factors = { 0, 7, 18, 99 } Company name = youtube URL https://www.youtube.com/watch?v=tVv6SrI9pU
27 Idx 11 Total Score = 166.0 Four Factors = { 0, 8, 34, 48 } Company name = com/ URL http://rockchurchsunnysvale.com/
28 Idx 9 Total Score = 154.0 Four Factors = { 0, 22, 23, 41 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ieep
29 Idx 17 Total Score = 77.0 Four Factors = { 0, 14, 8, 25 } Company name = youtube URL https://www.youtube.com/watch?v=1w70qIMMRc4
30 Idx 24 Total Score = 34.0 Four Factors = { 0, 7, 5, 5 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2012
```

### Insert rank 28 URL: [www.goofy.com](http://www.goofy.com)

```
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
2
Enter a Page Rank:
28
Enter a URL:
www.goofy.com

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
4
1 Idx 19 Total Score = 2132.0 Four Factors = { 364, 47, 69, 11 } Company name = thefreedictionary URL https://www.thefreedictionary.com/rock
2 Idx 15 Total Score = 512.0 Four Factors = { 69, 47, 14, 31 } Company name = merriam-webster URL https://www.merriam-webster.com/dictionary/rock
3 Idx 5 Total Score = 384.0 Four Factors = { 28, 16, 51, 59 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Rock\_\(geology\)
4 Idx 8 Total Score = 328.0 Four Factors = { 2, 23, 59, 95 } Company name = youtube URL https://www.youtube.com/watch?v=htlnQrFL5A
6 Idx 0 Total Score = 310.0 Four Factors = { 16, 8, 56, 46 } Company name = bing URL https://www.bing.com/search?q=rock
7 Idx 0 Total Score = 0.0 Four Factors = { 0, 0, 0, 0 } Company name = cat URL www.cat.com
7 Idx 10 Total Score = 308.0 Four Factors = { 0, 20, 57, 97 } Company name = foxnews URL https://www.foxnews.com/entertainment/jack-black-reunites-with
9 Idx 20 Total Score = 300.0 Four Factors = { 0, 14, 58, 98 } Company name = bing URL http://help.bing.microsoft.com/
10 Idx 23 Total Score = 292.0 Four Factors = { 0, 50, 38, 78 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2013
11 Idx 27 Total Score = 277.0 Four Factors = { 0, 24, 71, 16 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1963
12 Idx 13 Total Score = 269.0 Four Factors = { 0, 19, 68, 27 } Company name = youtube URL https://www.youtube.com/watch?v=9XaS93wMR0Q
13 Idx 18 Total Score = 262.0 Four Factors = { 0, 28, 66, 8 } Company name = rockauto URL https://www.rockauto.com/en/catalog/dodge
14 Idx 6 Total Score = 255.0 Four Factors = { 14, 36, 36, 5 } Company name = yelp URL https://www.yelp.com/biz/rok-bistro-sunnysvale-2
15 Idx 21 Total Score = 253.0 Four Factors = { 0, 15, 68, 19 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2011
16 Idx 1 Total Score = 250.0 Four Factors = { 0, 13, 57, 53 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford
17 Idx 22 Total Score = 250.0 Four Factors = { 0, 1, 52, 92 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2010
18 Idx 26 Total Score = 249.0 Four Factors = { 0, 30, 46, 51 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2019
19 Idx 4 Total Score = 249.0 Four Factors = { 0, 47, 25, 80 } Company name = rockauto URL https://www.rockauto.com/en/catalog/toyota
20 Idx 14 Total Score = 222.0 Four Factors = { 0, 10, 52, 46 } Company name = rockauto URL https://www.rockauto.com/
21 Idx 3 Total Score = 217.0 Four Factors = { 0, 2, 50, 63 } Company name = rockauto URL https://www.rockauto.com/en/catalog/
22 Idx 16 Total Score = 205.0 Four Factors = { 0, 16, 31, 80 } Company name = youtube URL https://www.youtube.com/watch?v=8SbUC-UaXfE
23 Idx 29 Total Score = 190.0 Four Factors = { 0, 40, 28, 26 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,1964
24 Idx 12 Total Score = 188.0 Four Factors = { 2, 20, 15, 93 } Company name = latimes URL http://www.latimes.com/books/la-ca-ic-jeff-tweedy-20181116-st
25 Idx 28 Total Score = 175.0 Four Factors = { 0, 11, 37, 42 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2018
26 Idx 2 Total Score = 167.0 Four Factors = { 0, 7, 18, 99 } Company name = youtube URL https://www.youtube.com/watch?v=tVv6SrI9pU
27 Idx 11 Total Score = 166.0 Four Factors = { 0, 8, 34, 48 } Company name = com/ URL http://rockchurchsunnysvale.com/
28 Idx 9 Total Score = 154.0 Four Factors = { 0, 22, 23, 41 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ieep
28 Idx 0 Total Score = 0.0 Four Factors = { 0, 0, 0, 0 } Company name = goofy URL www.goofy.com
29 Idx 17 Total Score = 77.0 Four Factors = { 0, 14, 8, 25 } Company name = youtube URL https://www.youtube.com/watch?v=1w70qIMMRc4
30 Idx 24 Total Score = 34.0 Four Factors = { 0, 7, 5, 5 } Company name = rockauto URL https://www.rockauto.com/en/catalog/ford,2012
```

## TEST BUCKET SORT: SORTING COMPANY NAMES:

Sorted based on Company names, which is the category before URLs:

Top first keyword cat:

The top first keyword is cat  
The results are sorted by company names

```
Idx 0 Total Score = 289.0 Four Factors = { 34, 24, 7, 50 } Company name = bing URL https://www.bing.com/search?q=cat
Idx 1 Total Score = 271.0 Four Factors = { 0, 25, 46, 83 } Company name = cat URL https://www.cat.com/
Idx 4 Total Score = 57.0 Four Factors = { 0, 21, 4, 3 } Company name = cat URL https://www.cat.com/en\_US/by-industry/mining/OperatorTraining.html
Idx 5 Total Score = 96.0 Four Factors = { 0, 1, 18, 40 } Company name = cat URL https://www.cat.com/en\_US/campaigns/awareness/shears-attachments.html
Idx 6 Total Score = 235.0 Four Factors = { 0, 48, 38, 25 } Company name = cat URL https://www.cat.com/en\_US/campaigns/awareness/attachments-2017-nat
Idx 12 Total Score = 266.0 Four Factors = { 0, 31, 59, 27 } Company name = cat URL https://parts.cat.com/en/catcorp/
Idx 15 Total Score = 217.0 Four Factors = { 0, 26, 36, 57 } Company name = cat URL https://www.cat.com/en\_US/campaigns/awareness/hammers-attachments
Idx 18 Total Score = 290.0 Four Factors = { 0, 3, 72, 68 } Company name = cat URL https://www.cat.com/en\_US/campaigns/awareness/cat-fluids.html?utm
Idx 19 Total Score = 137.0 Four Factors = { 0, 19, 23, 30 } Company name = cat URL https://www.cat.com/en\_US/campaigns/awareness/myaccount.html
Idx 20 Total Score = 274.0 Four Factors = { 0, 6, 59, 85 } Company name = cat URL https://www.cat.com/en\_US/campaigns/awareness/excavator-attachment
Idx 23 Total Score = 281.0 Four Factors = { 0, 39, 39, 86 } Company name = cat URL https://www.cat.com/fr\_US.html
Idx 24 Total Score = 301.0 Four Factors = { 0, 32, 77, 6 } Company name = cat URL https://www.cat.com/ko\_KR.html
Idx 26 Total Score = 217.0 Four Factors = { 0, 35, 25, 72 } Company name = cat URL https://www.cat.com/en\_ZA.html
Idx 27 Total Score = 78.0 Four Factors = { 0, 2, 8, 50 } Company name = cat URL https://www.cat.com/ru\_RU.html
Idx 28 Total Score = 300.0 Four Factors = { 0, 43, 71, 1 } Company name = cat URL https://www.cat.com/en\_MX.html
Idx 29 Total Score = 337.0 Four Factors = { 0, 43, 55, 86 } Company name = cat URL https://www.cat.com/tr\_TR.html
Idx 11 Total Score = 166.0 Four Factors = { 7, 4, 15, 78 } Company name = charlotteobserver URL https://www.charlotteobserver.com/news/state/south-ci
Idx 2 Total Score = 174.0 Four Factors = { 0, 43, 23, 19 } Company name = cnn URL https://money.cnn.com/quote/quote.html?symb=CAT
Idx 7 Total Score = 504.0 Four Factors = { 72, 41, 4, 50 } Company name = dictionary URL https://www.dictionary.com/browse/cat
Idx 25 Total Score = 157.0 Four Factors = { 0, 12, 35, 28 } Company name = googletagmanager URL https://www.googletagmanager.com/ns.html?id=GTM-5JB4
Idx 10 Total Score = 237.0 Four Factors = { 0, 18, 50, 51 } Company name = marketwatch URL https://www.marketwatch.com/tools/markets
Idx 16 Total Score = 253.0 Four Factors = { 0, 45, 27, 82 } Company name = marketwatch URL https://www.marketwatch.com/tools/markets/stocks
Idx 22 Total Score = 189.0 Four Factors = { 0, 40, 18, 55 } Company name = marketwatch URL https://www.marketwatch.com/tools/markets/stocks/country/
Idx 3 Total Score = 214.0 Four Factors = { 7, 39, 28, 17 } Company name = vetstreet URL http://www.vetstreet.com/cats/
Idx 21 Total Score = 1111.0 Four Factors = { 178, 38, 32, 49 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Cat
Idx 8 Total Score = 232.0 Four Factors = { 0, 32, 23, 99 } Company name = yahoo URL https://finance.yahoo.com/quote/CAT
Idx 9 Total Score = 165.0 Four Factors = { 1, 37, 8, 62 } Company name = youtube URL https://www.youtube.com/watch?v=yWiRvhfipI8
Idx 13 Total Score = 313.0 Four Factors = { 3, 14, 69, 63 } Company name = youtube URL https://www.youtube.com/watch?v=1wh8RzcQZr4
Idx 14 Total Score = 171.0 Four Factors = { 4, 6, 41, 16 } Company name = youtube URL https://www.youtube.com/watch?v=ppYts0nFOvs
Idx 17 Total Score = 113.0 Four Factors = { 3, 35, 1, 25 } Company name = youtube URL https://www.youtube.com/watch?v=jijMaW0ZGoA
```

## THE PROCEDURE OF HOW TO UNZIP FILES, INSTALL APPLICATION, AND RUN CODES

### PA1-YU-XIU.ZIP

PA1-Yu-Xiu/

**ProgramingHWReport.docx:** This is the assignment report.

**GoogleSimulator.jar:** This is the jar file of my GoogleSimulator project.

**GoogleSimulator/**

**src/**

**main/**

**java/**

GoogleSimulator.java

HeapSort.java

MyCrawler.java

WebPageInformation.java

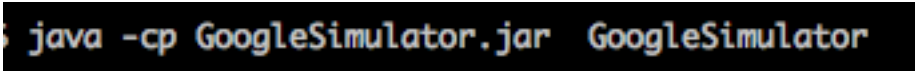
BinarySearchTree.java

BucketSort.java  
QuickSort.java  
KeyWordInformation.java  
META-INF/  
MANIFEST

### **RUN THE JAR FILE IN CONSOLE:**

java -cp GoogleSimulator.jar GoogleSimulator

**screen shot:**



```
java -cp GoogleSimulator.jar GoogleSimulator
```

### **PROCEDURE OF HOW TO RUN MY CODES:**

-----  
When running the file in console, user would see the command of

“Enter a number for a1”

“Enter a number for a2”

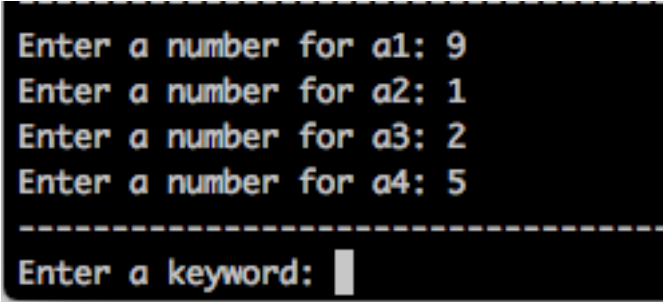
“Enter a number for a3”

“Enter a number for a4”

Inputting all the coefficient of the factors once, and it saves time of inputting those number each of the time when user type in a keyword. Recall the total score formula:

**Total Score** =  $a1 * \text{keyword frequency} + a2 * \text{web page exist age} + a3 * \text{number of outgoing links} + a4 * \text{advertisement payment}$

So that, even we use all same coefficient  $a1, a2, a3, a4$ , for all the keywords, we still have different total scores since the other four factors would be changed.



```
Enter a number for a1: 9
Enter a number for a2: 1
Enter a number for a3: 2
Enter a number for a4: 5
-----
Enter a keyword: 
```

After entering four numbers, user could enter a keyword and press enter, then the web crawler starts to crawl web pages. Example: “Enter a keyword: corgi”. It displays the



sorted 30 web pages with PageRank, Index, total score, four factors, company name, and URL:

```

PageRank 1 Idx 2 Total Score = 635.0 Four Factors = { 2, 46, 68, 87 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Welsh_corgi
PageRank 2 Idx 0 Total Score = 623.0 Four Factors = { 24, 19, 9, 74 } Company name = bing URL https://www.bing.com/search?q=corgi
PageRank 3 Idx 17 Total Score = 619.0 Four Factors = { 6, 40, 70, 77 } Company name = com/article/63484/11-short-facts-about-corgis URL http://mentalfloss.com/arti
acts-about-corgis
PageRank 4 Idx 14 Total Score = 596.0 Four Factors = { 5, 32, 57, 81 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Pembroke_Welsh_Corgi
PageRank 5 Idx 22 Total Score = 593.0 Four Factors = { 0, 37, 53, 90 } Company name = shacknews URL https://www.shacknews.com/article/101225/destiny-2-complete-str
PageRank 6 Idx 23 Total Score = 584.0 Four Factors = { 0, 34, 55, 88 } Company name = shacknews URL https://www.shacknews.com/article/108589/reggie-fils-aime-on-li
treaming-of-games-in-the-us
PageRank 7 Idx 12 Total Score = 581.0 Four Factors = { 3, 45, 27, 91 } Company name = nextdaypets URL https://www.nextdaypets.com/Corgi.htm
PageRank 8 Idx 19 Total Score = 579.0 Four Factors = { 0, 46, 49, 87 } Company name = ebay URL https://www.ebay.com/
PageRank 9 Idx 10 Total Score = 570.0 Four Factors = { 0, 25, 75, 79 } Company name = youtube URL https://www.youtube.com/watch?v=C0MJ2LHqbZ4
PageRank 10 Idx 5 Total Score = 568.0 Four Factors = { 2, 43, 71, 73 } Company name = corgi URL https://www.corgi.co.uk/
PageRank 11 Idx 21 Total Score = 565.0 Four Factors = { 0, 41, 52, 84 } Company name = shacknews URL https://www.shacknews.com/user/Shacknews/posts
PageRank 12 Idx 15 Total Score = 532.0 Four Factors = { 0, 4, 64, 80 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/breeds/
PageRank 13 Idx 4 Total Score = 510.0 Four Factors = { 0, 4, 38, 86 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Help:IPA/English
PageRank 14 Idx 27 Total Score = 488.0 Four Factors = { 0, 47, 13, 83 } Company name = shacknews URL https://www.shacknews.com/chatty?id=38076045
PageRank 15 Idx 8 Total Score = 457.0 Four Factors = { 21, 14, 17, 44 } Company name = com/corgi/ URL http://pluspets.com/corgi/
PageRank 16 Idx 6 Total Score = 454.0 Four Factors = { 0, 11, 14, 83 } Company name = youtube URL https://www.youtube.com/watch?v=2iNGMRRy6ls
PageRank 17 Idx 26 Total Score = 419.0 Four Factors = { 0, 30, 2, 77 } Company name = shacknews URL https://www.shacknews.com/topic/esports
PageRank 18 Idx 1 Total Score = 411.0 Four Factors = { 0, 47, 27, 62 } Company name = shacknews URL https://www.shacknews.com/article/108581/world-of-warcraft-cele
-corgi-goggles
PageRank 19 Idx 9 Total Score = 407.0 Four Factors = { 5, 9, 39, 55 } Company name = ebay URL https://www.ebay.com/sch/i.html?_nkw=corgi
PageRank 20 Idx 7 Total Score = 385.0 Four Factors = { 3, 12, 43, 52 } Company name = com/blog/2018/11/07/corgi-butt-heated-pillow/ URL https://technabob.com/blog/
t-heated-pillow/
PageRank 21 Idx 25 Total Score = 336.0 Four Factors = { 0, 26, 60, 38 } Company name = shacknews URL https://www.shacknews.com/article/108598/sega-drops-a-new-team
PageRank 22 Idx 16 Total Score = 329.0 Four Factors = { 1, 18, 51, 40 } Company name = youtube URL https://www.youtube.com/watch?v=nu77ETGoxAY
PageRank 23 Idx 11 Total Score = 279.0 Four Factors = { 0, 11, 59, 30 } Company name = akc URL https://www.akc.org/dog-breeds/pembroke-welsh-corgi/
PageRank 24 Idx 13 Total Score = 271.0 Four Factors = { 0, 1, 55, 32 } Company name = corgi-usa URL https://www.corgi-usa.com/
PageRank 25 Idx 20 Total Score = 270.0 Four Factors = { 0, 2, 14, 48 } Company name = shacknews URL https://www.shacknews.com/topic/fortnite
PageRank 26 Idx 24 Total Score = 261.0 Four Factors = { 0, 12, 12, 45 } Company name = shacknews URL https://www.shacknews.com/login/login
PageRank 27 Idx 3 Total Score = 227.0 Four Factors = { 0, 41, 13, 32 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/dogs/sale/
PageRank 28 Idx 29 Total Score = 223.0 Four Factors = { 0, 37, 38, 22 } Company name = shacknews URL https://www.shacknews.com/topic/monster-hunter-world
PageRank 29 Idx 28 Total Score = 135.0 Four Factors = { 0, 21, 47, 4 } Company name = shacknews URL https://www.shacknews.com/author/charles-singletary-jr
PageRank 30 Idx 18 Total Score = 104.0 Four Factors = { 0, 9, 40, 3 } Company name = com/dog-breeds/pembroke-welsh-corgi URL https://dogtime.com/dog-breeds/pembroke

```

At the end of the 30 web pages, the user would see a sentence: “Test Process BST? Y/N: ”. At this moment, user could type Y, which means Yes -> test BST; or put N, which means No -> not testing BST.

```

PageRank 30 Idx 18 Total Score = 104.0 F
Test Process BST? Y/N: 

```

Enter N: User could keep input keyword

```

Test Process BST? Y/N: N
-----
Enter a keyword: 

```

Enter Y: Test BST features

User would see an options bar: “Options: 1. Search Page Rank; 2. Insert URE; 3. Delete URL; 4. Show tree; 0. Exit;”

```
Test Process BST? Y/N: Y
```

```
-----  
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;  
█
```

User could choose a number to test the feature of the BST:

#### **Test Search a page rank:**

If user puts 1, which means the search page rank feature would be tested, and then enter a page rank, from 1 to 30, to search the rank. It would be displayed:

```
Page Rank 0 Idx 20 Total Score = 2077.0 Four Factors = { 0, 0, 10, 0 } Company name = con.org or con.org/pem/one.html con.org one.html  
Test Process BST? Y/N: Y  
-----  
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;  
1  
Enter a Page Rank:  
2  
2 Idx 0 Total Score = 623.0 Four Factors = { 24, 19, 9, 74 } Company name = bing URL https://www.bing.com/search?q=corgi
```

#### **Test Insertion of URL:**

The user would see the option bar again and choosing the feature wants to test:

Enter: 2

Enter a Page Rank:

3

Enter a URL:

[www.google.com](http://www.google.com)

then, **choose 4** in option bar to show the tree after insert [www.google.com](http://www.google.com)  
this URL would be inserted and with a rank 3;

```

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
2
Enter a Page Rank:
3
Enter a URL:
www.google.com
-----
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
4
1 Idx 2 Total Score = 635.0 Four Factors = { 2, 46, 68, 87 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Welsh_corgi
2 Idx 0 Total Score = 623.0 Four Factors = { 24, 19, 9, 74 } Company name = bing URL https://www.bing.com/search?q=corgi
3 Idx 17 Total Score = 619.0 Four Factors = { 6, 40, 70, 77 } Company name = com/article/63484/11-short-facts-about-corgis URL http://mentalfloss.com/ar
t-corgis
3 Idx 0 Total Score = 0.0 Four Factors = { 0, 0, 0, 0 } Company name = google URL www.google.com
4 Idx 14 Total Score = 596.0 Four Factors = { 5, 32, 57, 81 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Pembroke_Welsh_Corgi
5 Idx 22 Total Score = 593.0 Four Factors = { 0, 37, 53, 90 } Company name = shacknews URL https://www.shacknews.com/article/101225/destiny-2-complete-s
6 Idx 23 Total Score = 584.0 Four Factors = { 0, 34, 55, 88 } Company name = shacknews URL https://www.shacknews.com/article/108589/reggie-fils-aime-on-
of-games-in-the-us
7 Idx 12 Total Score = 581.0 Four Factors = { 3, 45, 27, 91 } Company name = nextdaypets URL https://www.nextdaypets.com/Corgi.htm
8 Idx 19 Total Score = 579.0 Four Factors = { 0, 46, 49, 87 } Company name = ebay URL https://www.ebay.com/
9 Idx 10 Total Score = 570.0 Four Factors = { 0, 25, 75, 79 } Company name = youtube URL https://www.youtube.com/watch?v=C0MJ2LHqbZ4
10 Idx 5 Total Score = 568.0 Four Factors = { 2, 43, 71, 73 } Company name = corgi URL https://www.corgi.co.uk/
11 Idx 21 Total Score = 565.0 Four Factors = { 0, 41, 52, 84 } Company name = shacknews URL https://www.shacknews.com/user/Shacknews/posts
12 Idx 15 Total Score = 532.0 Four Factors = { 0, 4, 64, 80 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/breeds/
13 Idx 4 Total Score = 510.0 Four Factors = { 0, 4, 38, 86 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Help:IPA/English
14 Idx 27 Total Score = 488.0 Four Factors = { 0, 47, 13, 83 } Company name = shacknews URL https://www.shacknews.com/chatty?id=38076045
15 Idx 8 Total Score = 457.0 Four Factors = { 21, 14, 17, 44 } Company name = com/corgi/ URL http://pluspets.com/corgi/
16 Idx 6 Total Score = 454.0 Four Factors = { 0, 11, 14, 83 } Company name = youtube URL https://www.youtube.com/watch?v=2iN6MRRy6ls
17 Idx 26 Total Score = 419.0 Four Factors = { 0, 30, 2, 77 } Company name = shacknews URL https://www.shacknews.com/topic/esports
18 Idx 1 Total Score = 411.0 Four Factors = { 0, 47, 27, 62 } Company name = shacknews URL https://www.shacknews.com/article/108581/world-of-warcraft-ce
ggles
19 Idx 9 Total Score = 407.0 Four Factors = { 5, 9, 39, 55 } Company name = ebay URL https://www.ebay.com/sch/i.html?_nkw=corgi
20 Idx 7 Total Score = 385.0 Four Factors = { 3, 12, 43, 52 } Company name = com/blog/2018/11/07/corgi-butt-heated-pillow/ URL https://technabob.com/blo
pillow/
21 Idx 25 Total Score = 336.0 Four Factors = { 0, 26, 60, 38 } Company name = shacknews URL https://www.shacknews.com/article/108598/sega-drops-a-new-te
22 Idx 16 Total Score = 329.0 Four Factors = { 1, 18, 51, 40 } Company name = youtube URL https://www.youtube.com/watch?v=nu77ETGoxAY
23 Idx 11 Total Score = 279.0 Four Factors = { 0, 11, 59, 30 } Company name = akc URL https://www.akc.org/dog-breeds/pembroke-welsh-corgi/
24 Idx 13 Total Score = 271.0 Four Factors = { 0, 1, 55, 32 } Company name = corgi-usa URL https://www.corgi-usa.com/
25 Idx 20 Total Score = 270.0 Four Factors = { 0, 2, 14, 48 } Company name = shacknews URL https://www.shacknews.com/topic/fortnite
26 Idx 24 Total Score = 261.0 Four Factors = { 0, 12, 12, 45 } Company name = shacknews URL https://www.shacknews.com/login/login
27 Idx 3 Total Score = 227.0 Four Factors = { 0, 41, 13, 32 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/dogs/sale/

28 Idx 29 Total Score = 223.0 Four Factors = { 0, 37, 38, 22 } Company name = shacknews URL https://www.shacknews.com/topic/monster-hunter-world
29 Idx 28 Total Score = 135.0 Four Factors = { 0, 21, 47, 4 } Company name = shacknews URL https://www.shacknews.com/author/charles-singletary-jr
30 Idx 18 Total Score = 104.0 Four Factors = { 0, 9, 40, 3 } Company name = com/dog-breeds/pembroke-welsh-corgi URL https://dogtime.com/dog-breeds/pembroke-welsh-corgi
-----
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;

```

## Testing deletion of an URL:

Choose option 3;

Enter a page rank: 5

Choose option 4 to display the tree after deleting URL of rank 5

```

Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
3
Enter a Page Rank:
5
-----
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
4
1 Idx 2 Total Score = 635.0 Four Factors = { 2, 46, 68, 87 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Welsh_Corgi
2 Idx 0 Total Score = 623.0 Four Factors = { 24, 19, 9, 74 } Company name = bing URL https://www.bing.com/search?q=corgi
3 Idx 17 Total Score = 619.0 Four Factors = { 6, 40, 70, 77 } Company name = com/article/63484/11-short-facts-about-corgis URL http://mentalfloss.com/article/63484/11-short-facts-about-corgis
3 Idx 0 Total Score = 0.0 Four Factors = { 0, 0, 0, 0 } Company name = google URL www.google.com
4 Idx 14 Total Score = 596.0 Four Factors = { 5, 32, 57, 81 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Pembroke_Welsh_Corgi
6 Idx 23 Total Score = 584.0 Four Factors = { 0, 34, 55, 88 } Company name = shacknews URL https://www.shacknews.com/article/108589/reggie-fils-aime-on-limiting-cloud-ba
of-games-in-the-us
7 Idx 12 Total Score = 581.0 Four Factors = { 3, 45, 27, 91 } Company name = nextdaypets URL https://www.nextdaypets.com/Corgi.htm
8 Idx 19 Total Score = 579.0 Four Factors = { 0, 46, 49, 87 } Company name = ebay URL https://www.ebay.com/
9 Idx 10 Total Score = 570.0 Four Factors = { 0, 25, 75, 79 } Company name = youtube URL https://www.youtube.com/watch?v=C0MJ2LHqbZ4
10 Idx 5 Total Score = 568.0 Four Factors = { 2, 43, 71, 73 } Company name = corgi URL https://www.corgi.co.uk/
11 Idx 21 Total Score = 565.0 Four Factors = { 0, 41, 52, 84 } Company name = shacknews URL https://www.shacknews.com/user/Shacknews/posts
12 Idx 15 Total Score = 532.0 Four Factors = { 0, 4, 64, 80 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/breeds/
13 Idx 4 Total Score = 510.0 Four Factors = { 0, 4, 38, 86 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Help:IPA/English
14 Idx 27 Total Score = 488.0 Four Factors = { 0, 47, 13, 83 } Company name = shacknews URL https://www.shacknews.com/chatty?id=38076045
15 Idx 8 Total Score = 457.0 Four Factors = { 21, 14, 17, 44 } Company name = com/corgi/ URL http://pluspets.com/corgi/
16 Idx 6 Total Score = 454.0 Four Factors = { 0, 11, 14, 83 } Company name = youtube URL https://www.youtube.com/watch?v=2iNGMRRy6ls
17 Idx 26 Total Score = 419.0 Four Factors = { 0, 30, 2, 77 } Company name = shacknews URL https://www.shacknews.com/topic/esports
18 Idx 1 Total Score = 411.0 Four Factors = { 0, 47, 27, 62 } Company name = shacknews URL https://www.shacknews.com/article/108581/world-of-warcraft-celebrates-14-years
gales
19 Idx 9 Total Score = 407.0 Four Factors = { 5, 9, 39, 55 } Company name = ebay URL https://www.ebay.com/sch/i.html?_nkw=corgi
20 Idx 7 Total Score = 385.0 Four Factors = { 3, 12, 43, 52 } Company name = com/blog/2018/11/07/corgi-butt-heated-pillow/ URL https://technabob.com/blog/2018/11/07/corgi
pillow/
21 Idx 25 Total Score = 336.0 Four Factors = { 0, 26, 60, 38 } Company name = shacknews URL https://www.shacknews.com/article/108598/sega-drops-a-new-team-sonic-racing-b
22 Idx 16 Total Score = 329.0 Four Factors = { 1, 18, 51, 40 } Company name = youtube URL https://www.youtube.com/watch?v=nu77ETGoxAY
23 Idx 11 Total Score = 279.0 Four Factors = { 0, 11, 59, 30 } Company name = akc URL https://www.akc.org/dog-breeds/pembroke-welsh-corgi/
24 Idx 13 Total Score = 271.0 Four Factors = { 0, 1, 55, 32 } Company name = corgi-usa URL https://www.corgi-usa.com/
25 Idx 20 Total Score = 270.0 Four Factors = { 0, 2, 14, 48 } Company name = shacknews URL https://www.shacknews.com/topic/fortnite
26 Idx 24 Total Score = 261.0 Four Factors = { 0, 12, 12, 45 } Company name = shacknews URL https://www.shacknews.com/login/login
27 Idx 3 Total Score = 227.0 Four Factors = { 0, 41, 13, 32 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/dogs/sale/
28 Idx 29 Total Score = 223.0 Four Factors = { 0, 37, 38, 22 } Company name = shacknews URL https://www.shacknews.com/topic/monster-hunter-world
29 Idx 28 Total Score = 135.0 Four Factors = { 0, 21, 47, 4 } Company name = shacknews URL https://www.shacknews.com/author/charles-singletary-jr
30 Idx 18 Total Score = 104.0 Four Factors = { 0, 9, 40, 3 } Company name = com/dog-breeds/pembroke-welsh-corgi URL https://dogtime.com/dog-breeds/pembroke-welsh-corgi

```

Exist the test BST after testing searching, deletion and insertion:

Enter: 0

Then user can input another keyword: cat

```

-----
Options: 1. Search Page Rank; 2. Insert URL; 3. Delete URL; 4. Show tree; 0. Exit;
0
-----
Enter a keyword: cat

```



Crawl web page again:

```
PageRank 1 Idx 21 Total Score = 1744.0 Four Factors = { 178, 8, 22, 18 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Cat
PageRank 2 Idx 7 Total Score = 696.0 Four Factors = { 72, 6, 16, 2 } Company name = dictionary URL https://www.dictionary.com/browse/cat
PageRank 3 Idx 3 Total Score = 676.0 Four Factors = { 7, 7, 53, 100 } Company name = vetstreet URL http://www.vetstreet.com/cats/
PageRank 4 Idx 6 Total Score = 602.0 Four Factors = { 0, 49, 79, 79 } Company name = cat URL https://www.cat.com/en_US/campaigns/awareness/attachments-2017-national-offer.html
PageRank 5 Idx 1 Total Score = 601.0 Four Factors = { 0, 50, 28, 99 } Company name = cat URL https://www.cat.com/
PageRank 6 Idx 2 Total Score = 565.0 Four Factors = { 0, 15, 30, 98 } Company name = cnn URL https://money.cnn.com/quote/quote.html?symb=CAT
PageRank 7 Idx 9 Total Score = 553.0 Four Factors = { 2, 14, 53, 83 } Company name = youtube URL https://www.youtube.com/watch?v=yWjRvhfjpI8
PageRank 8 Idx 16 Total Score = 524.0 Four Factors = { 0, 30, 22, 90 } Company name = marketwatch URL https://www.marketwatch.com/tools/markets/stocks
PageRank 9 Idx 15 Total Score = 515.0 Four Factors = { 0, 20, 60, 75 } Company name = cat URL https://www.cat.com/en_US/campaigns/awareness/hammers-attachments.html
PageRank 10 Idx 27 Total Score = 490.0 Four Factors = { 0, 30, 35, 78 } Company name = cat URL https://www.cat.com/ru_RU.html
PageRank 11 Idx 0 Total Score = 486.0 Four Factors = { 34, 48, 6, 24 } Company name = bing URL https://www.bing.com/search?q=cat
PageRank 12 Idx 12 Total Score = 483.0 Four Factors = { 0, 1, 51, 76 } Company name = cat URL https://parts.cat.com/en/catcorp/
PageRank 13 Idx 24 Total Score = 454.0 Four Factors = { 0, 8, 68, 62 } Company name = cat URL https://www.cat.com/ko_KR.html
PageRank 14 Idx 14 Total Score = 451.0 Four Factors = { 6, 17, 50, 56 } Company name = youtube URL https://www.youtube.com/watch?v=opYts0nF0vs
PageRank 15 Idx 13 Total Score = 428.0 Four Factors = { 3, 11, 40, 62 } Company name = youtube URL https://www.youtube.com/watch?v=1Wh8RzcQZr4
PageRank 16 Idx 17 Total Score = 398.0 Four Factors = { 3, 6, 50, 53 } Company name = youtube URL https://www.youtube.com/watch?v=JiyMdlW0ZGoA
PageRank 17 Idx 18 Total Score = 381.0 Four Factors = { 0, 26, 55, 49 } Company name = cat URL https://www.cat.com/en_US/campaigns/awareness/cat-fluids.html?utm_content=ps_anal_fluids&utm_source=coegi&utm_medium=video&utm_campaign=video_15s&utm_term=yoga_testimonial
PageRank 18 Idx 23 Total Score = 365.0 Four Factors = { 0, 10, 15, 65 } Company name = cat URL https://www.cat.com/fr_US.html
PageRank 19 Idx 8 Total Score = 363.0 Four Factors = { 0, 46, 31, 51 } Company name = yahoo URL https://finance.yahoo.com/quote/CAT
PageRank 20 Idx 22 Total Score = 356.0 Four Factors = { 0, 14, 46, 50 } Company name = marketwatch URL https://www.marketwatch.com/tools/markets/stocks/country/united-states
PageRank 21 Idx 11 Total Score = 329.0 Four Factors = { 7, 43, 39, 29 } Company name = charlotteobserver URL https://www.charlotteobserver.com/news/state/south-carolina/article65.html
PageRank 22 Idx 26 Total Score = 273.0 Four Factors = { 0, 34, 57, 25 } Company name = cat URL https://www.cat.com/en_ZA.html
PageRank 23 Idx 4 Total Score = 256.0 Four Factors = { 0, 31, 55, 23 } Company name = cat URL https://www.cat.com/en_US/by-industry/mining/OperatorTraining.html
PageRank 24 Idx 5 Total Score = 234.0 Four Factors = { 0, 4, 10, 42 } Company name = cat URL https://www.cat.com/en_US/campaigns/awareness/shears-attachments.html?utm_content=ls_an_educational_northamerica_gci&utm_source=google&utm_medium=cpc&utm_campaign=united_states_-_shears-cn-wt-p&utm_term=catwt66763-shears-728x90
PageRank 25 Idx 20 Total Score = 159.0 Four Factors = { 0, 23, 53, 6 } Company name = cat URL https://www.cat.com/en_US/campaigns/awareness/excavator-attachments.html?utm_content=tools_an_educational_northamerica_gci&utm_source=google&utm_medium=cpc&utm_campaign=united_states_-_demomp-cn-wt-p&utm_term=catwt66763-generic-728x90
PageRank 26 Idx 10 Total Score = 159.0 Four Factors = { 0, 15, 42, 12 } Company name = marketwatch URL https://www.marketwatch.com/tools/markets
PageRank 27 Idx 29 Total Score = 155.0 Four Factors = { 0, 22, 59, 3 } Company name = cat URL https://www.cat.com/tr_TR.html
PageRank 28 Idx 25 Total Score = 154.0 Four Factors = { 0, 30, 12, 20 } Company name = googletagmanager URL https://www.googletagmanager.com/ns.html?id=GTM-5J84X4
PageRank 29 Idx 19 Total Score = 111.0 Four Factors = { 0, 15, 3, 18 } Company name = cat URL https://www.cat.com/en_US/campaigns/awareness/myaccount.html
PageRank 30 Idx 28 Total Score = 80.0 Four Factors = { 0, 47, 14, 1 } Company name = cat URL https://www.cat.com/en_MX.html
```

Enter: N

Enter keyword corgi again, and crawl web pages

Test Process BST? Y/N: N

Enter a keyword:

```

PageRank 1 Idx 0 Total Score = 711.0 Four Factors = { 17, 12, 43, 92 } Company name = bing URL https://www.bing.com/search?q=corgi
PageRank 2 Idx 15 Total Score = 617.0 Four Factors = { 0, 34, 69, 89 } Company name = forpaws URL http://www.forpaws.org/contact/index.html
PageRank 3 Idx 21 Total Score = 610.0 Four Factors = { 0, 47, 59, 89 } Company name = bellaonline URL http://www.bellaonline.com/subjects/557.asp
PageRank 4 Idx 14 Total Score = 570.0 Four Factors = { 0, 41, 37, 91 } Company name = forpaws URL http://www.forpaws.org/adopt/index.html
PageRank 5 Idx 27 Total Score = 546.0 Four Factors = { 0, 14, 51, 86 } Company name = shacknews URL https://www.shacknews.com/article/108598/sega-drops-a-new-team-sonic-
PageRank 6 Idx 5 Total Score = 535.0 Four Factors = { 3, 9, 12, 95 } Company name = com/blog/2018/11/07/corgi-butt-heated-pillow/ URL https://technabob.com/blog/2018/11/
heated-pillow/
PageRank 7 Idx 25 Total Score = 522.0 Four Factors = { 0, 4, 54, 82 } Company name = shacknews URL https://www.shacknews.com/article/108589/reggie-fils-aime-on-limiting-
reaming-of-games-in-the-us
PageRank 8 Idx 24 Total Score = 505.0 Four Factors = { 0, 19, 78, 66 } Company name = shacknews URL https://www.shacknews.com/article/101225/destiny-2-complete-strategy-
PageRank 9 Idx 1 Total Score = 491.0 Four Factors = { 4, 19, 58, 64 } Company name = forpaws URL http://www.forpaws.org/articles/first-time-owner.htm
PageRank 10 Idx 17 Total Score = 481.0 Four Factors = { 0, 23, 69, 64 } Company name = forpaws URL http://www.forpaws.org/articles/alpha.htm
PageRank 11 Idx 6 Total Score = 474.0 Four Factors = { 0, 6, 34, 80 } Company name = com/OhMyCorgi/status/1062873383443013633 URL https://twitter.com/OhMyCorgi/status/10
633
PageRank 12 Idx 8 Total Score = 425.0 Four Factors = { 0, 34, 33, 65 } Company name = com/OhMyCorgi/status/1062000829568299013 URL https://twitter.com/OhMyCorgi/status/1
9013
PageRank 13 Idx 19 Total Score = 424.0 Four Factors = { 1, 17, 39, 64 } Company name = forpaws URL http://www.forpaws.org/articles/index2.html
PageRank 14 Idx 18 Total Score = 423.0 Four Factors = { 0, 39, 52, 56 } Company name = forpaws URL http://www.forpaws.org/articles/children2.htm
PageRank 15 Idx 22 Total Score = 416.0 Four Factors = { 0, 18, 79, 48 } Company name = shacknews URL https://www.shacknews.com/topic/fornite
PageRank 16 Idx 2 Total Score = 413.0 Four Factors = { 0, 25, 19, 70 } Company name = shacknews URL https://www.shacknews.com/article/108581/world-of-warcraft-celebrates
-corgi-goggles
PageRank 17 Idx 7 Total Score = 384.0 Four Factors = { 0, 49, 55, 45 } Company name = com/OhMyCorgi/status/1062363249105166336 URL https://twitter.com/OhMyCorgi/status/1
6336
PageRank 18 Idx 13 Total Score = 362.0 Four Factors = { 0, 27, 75, 37 } Company name = com/dog-breeds/pembroke-welsh-corgi URL https://dogtime.com/dog-breeds/pembroke-we
PageRank 19 Idx 9 Total Score = 289.0 Four Factors = { 0, 29, 5, 50 } Company name = akc URL https://www.akc.org/dog-breeds/pembroke-welsh-corgi/
PageRank 20 Idx 12 Total Score = 287.0 Four Factors = { 6, 41, 71, 10 } Company name = com/article/63484/11-short-facts-about-corgis URL http://mentalfloss.com/article/6
facts-about-corgis
PageRank 21 Idx 23 Total Score = 271.0 Four Factors = { 0, 40, 58, 23 } Company name = shacknews URL https://www.shacknews.com/user/Shacknews/posts
PageRank 22 Idx 20 Total Score = 268.0 Four Factors = { 0, 46, 41, 28 } Company name = forpaws URL http://www.forpaws.org/
PageRank 23 Idx 3 Total Score = 261.0 Four Factors = { 2, 48, 40, 23 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Welsh_corgi
PageRank 24 Idx 10 Total Score = 247.0 Four Factors = { 5, 32, 30, 22 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Pembroke_Welsh_Corgi
PageRank 25 Idx 28 Total Score = 235.0 Four Factors = { 0, 26, 32, 29 } Company name = shacknews URL https://www.shacknews.com/topic/esports
PageRank 26 Idx 16 Total Score = 232.0 Four Factors = { 0, 7, 30, 33 } Company name = selectsmart URL http://www.selectsmart.com/DOG/
PageRank 27 Idx 26 Total Score = 217.0 Four Factors = { 0, 27, 70, 10 } Company name = shacknews URL https://www.shacknews.com/login/login
PageRank 28 Idx 29 Total Score = 212.0 Four Factors = { 0, 18, 12, 34 } Company name = shacknews URL https://www.shacknews.com/chatty?id=38076045
PageRank 29 Idx 4 Total Score = 188.0 Four Factors = { 0, 18, 75, 4 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Help:IPA/English
PageRank 30 Idx 11 Total Score = 113.0 Four Factors = { 0, 35, 4, 14 } Company name = reference URL https://www.reference.com/pets-animals/mini-corgi-337975f863fc2a8c
Test Process BST? Y/N: ☐

```

The user can keep entering keyword by enter N after display 30 pages;

```
Test Process BST? Y/N: N
```

```
-----
Enter a keyword: 
```

At any time, user can input an empty keyword to exit the loop. Then the program would show the most popular words based the number of keywords entered.

Bucket sort would sort the top first popular keyword's 30 company names:

```
Test Process BS? Y/N: N
-----
Enter a keyword:
Count 2 Keyword corgi
Count 1 Keyword cat
The top first keyword is corgi
The results are sorted by company names
-----
Idx 2 Total Score = 394.0 Four Factors = { 0, 26, 26, 58 } Company name = adoptapet URL https://www.adoptapet.com/s/adopt-a-corgi
Idx 9 Total Score = 670.0 Four Factors = { 0, 28, 49, 99 } Company name = akc URL https://www.akc.org/dog-breeds/pembroke-welsh-corgi/
Idx 0 Total Score = 720.0 Four Factors = { 16, 2, 42, 96 } Company name = bing URL https://www.bing.com/search?q=corgi
Idx 8 Total Score = 84.0 Four Factors = { 0, 1, 1, 16 } Company name = com/0hMyCorgi/status/1062363249105166336 URL https://twitter.com/0hMyCorgi/s
Idx 6 Total Score = 121.0 Four Factors = { 0, 8, 21, 10 } Company name = com/0hMyCorgi/status/1062873383443013633 URL https://twitter.com/0hMyCorgi
Idx 12 Total Score = 266.0 Four Factors = { 6, 24, 10, 34 } Company name = com/article/63484/11-short-facts-about-corgis URL http://mentalfloss.com/corgis
Idx 5 Total Score = 546.0 Four Factors = { 3, 30, 55, 66 } Company name = com/blog/2018/11/07/corgi-butt-heated-pillow/ URL https://technabob.com/blog/
Idx 13 Total Score = 529.0 Four Factors = { 0, 32, 4, 97 } Company name = com/dog-breeds/pembroke-welsh-corgi URL https://dogtime.com/dog-breeds/pembroke-welsh-corgi/
Idx 25 Total Score = 409.0 Four Factors = { 0, 32, 74, 31 } Company name = com/en-us/news/22770319/celebrate-14-years-of-world-of-warcraft URL http://www.1up.com/news/22770319/celebrate-14-years-of-world-of-warcraft
Idx 24 Total Score = 582.0 Four Factors = { 0, 27, 45, 84 } Company name = facebook URL https://www.facebook.com/shacknews
Idx 4 Total Score = 231.0 Four Factors = { 0, 33, 1, 39 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/dogs/sale/
Idx 10 Total Score = 183.0 Four Factors = { 3, 11, 2, 29 } Company name = nextdaypets URL https://www.nextdaypets.com/Corgi.htm
Idx 11 Total Score = 424.0 Four Factors = { 0, 50, 23, 61 } Company name = nextdaypets URL https://www.nextdaypets.com/directory/breeds/
Idx 1 Total Score = 619.0 Four Factors = { 0, 44, 35, 94 } Company name = shacknews URL https://www.shacknews.com/article/108581/world-of-warcraft-esports
Idx 14 Total Score = 289.0 Four Factors = { 0, 8, 2, 55 } Company name = shacknews URL https://www.shacknews.com/topic/fortnite
Idx 15 Total Score = 498.0 Four Factors = { 0, 20, 56, 62 } Company name = shacknews URL https://www.shacknews.com/user/Shacknews/posts
Idx 16 Total Score = 365.0 Four Factors = { 0, 8, 9, 66 } Company name = shacknews URL https://www.shacknews.com/article/101225/destiny-2-complete-guide
Idx 17 Total Score = 156.0 Four Factors = { 0, 27, 13, 18 } Company name = shacknews URL https://www.shacknews.com/article/108589/reggie-filsaime-games-in-the-us
Idx 18 Total Score = 318.0 Four Factors = { 0, 24, 33, 39 } Company name = shacknews URL https://www.shacknews.com/login/login
Idx 19 Total Score = 193.0 Four Factors = { 0, 17, 2, 34 } Company name = shacknews URL https://www.shacknews.com/article/108598/segas-drops-a-new-trophy
Idx 20 Total Score = 741.0 Four Factors = { 0, 41, 80, 92 } Company name = shacknews URL https://www.shacknews.com/topic/esports
Idx 21 Total Score = 323.0 Four Factors = { 0, 11, 4, 60 } Company name = shacknews URL https://www.shacknews.com/chatty?id=38076045
Idx 22 Total Score = 648.0 Four Factors = { 0, 39, 63, 84 } Company name = shacknews URL https://www.shacknews.com/author/charles-singletary-jr
Idx 23 Total Score = 514.0 Four Factors = { 0, 47, 59, 58 } Company name = shacknews URL https://www.shacknews.com/topic/monster-hunter-world
Idx 26 Total Score = 558.0 Four Factors = { 0, 36, 24, 90 } Company name = shacknews URL https://www.shacknews.com/login/register
Idx 28 Total Score = 192.0 Four Factors = { 0, 14, 11, 29 } Company name = shacknews URL https://www.shacknews.com/article/102661/video-game-releases-2018
Idx 29 Total Score = 615.0 Four Factors = { 0, 11, 38, 98 } Company name = shacknews URL https://www.shacknews.com/article/108590/final-fantasy-7-rebirth-tetsuya-nomura
Idx 3 Total Score = 505.0 Four Factors = { 2, 24, 49, 64 } Company name = wikipedia URL https://en.wikipedia.org/wiki/Welsh_corgi
Idx 7 Total Score = 281.0 Four Factors = { 1, 1, 61, 18 } Company name = youtube URL https://www.youtube.com/watch?v=iMBBtU3DzP4
Idx 27 Total Score = 212.0 Four Factors = { 0, 19, 61, 2 } Company name = youtube URL https://www.youtube.com/user/Shacknewsgames
```

## PROBLEMS ENCOUNTERED DURING THE IMPLEMENTATION.

1. My quick sort partition function cannot work correctly at the beginning, then I figured out that I made a mistake in editing a condition.
2. After changing the filter, my modified web crawler cannot give me 30 web pages.
3. It took me a while to figure out how to split an URL into several strings and insert a company's name into a bucket.
4. Bucket sort is hard to build since the pseudocode is not that detailed like other sorting algorithm. It is also hard to make the insertion sort work correctly.
5. It took me a long time to make BST work. The arguments and function calls, and the design of how and when let user to choose to delete and insert URL is also hard.

6. I hope to implement a GUI for this project. I created a single frame, but I finally focused on the console version.

## LESSONS LEARNED

1. In this programming assignment, I learned how to implement quick sort to sort all the web page information by comparing each web page's score.
2. I learned how to modify my web crawler, crawler4j, into a keyword based web crawler instead of the web pages/ links based web crawler.
3. I learned how to implement bucket sort with insertion sort to sort companies' names.
4. I learned how to split URL and put the correct company name into a bucket.
5. I learned how to create a content page in word.
6. I should start even more earlier to do this assignment, then I could have more time to make it better.