Search Engine Project

Youssef Elmahdy - 900212370

Instructor: Dr. Mohamed Alhalaby

CSCE220304 – Analysis and Design of Algorithms Lab

The American University in Cairo

a) Pseudo-code for the indexing and ranking algorithms

b) Time and space complexity analysis for your indexing and ranking algorithms

**Indexing Algorithm:**

Pseudo-code:

while(true)

{

while(not end of Test1)

{

initializing data for each line in string stream

splitting data in the stram with delimiter ','

if input query is equal to data from index 1 to size of array - 1

then push data of index 0 in the array to the vector

increase p by one

}

for(i = 0 to p)

{

while(not end of Test2)

{

initializing data for each line in string stream

splitting data in the stram with delimiter ','

if data is equal to data with index 1

then calculate the rank and then push it into the vector

}

}

}

Time complexity = N \* N \* N \* N = O(N^4).

Space complexity = O(1) + O(1) + O(1) + O(1) = O(1).

**Ranking Algorithm:**

Pseudo-code:

while(not end of Test)

{

initializing data for each line in string stream

splitting data in the stram with delimiter ','

while(not end of stream)

{

if data is equal to data with index 1 or 0

then increment the impressions to calculate the page rank

then push it in the vector

}

after selection of page to view the CTR is incremented and then recalculate the page rank using the formula

update score in text file

}

Complexity:

Time complexity = N \* N = O(N^2).

Space complexity = O(1) + O(1) = O(1).

c) The main data structures used by your algorithm

I have used the following:

1. Vector of pairs
2. Vector of string
3. Vector of double
4. String data
5. Int data

d) Any design tradeoffs you made along with their justifications

1. I’ve opted for the use of a vector of pairs because vectors have a much better time and space complexity than a map data structure, which would’ve been easier to implement. But, the vector has a better time and space complexity of O(1) for storing and accessing data. While, in a map, inserting and accessing data takes O(log N) time.
2. I’ve used a vector of pairs to store link between each website and the other, rather than using a vector of vectors, which would’ve been easier to implement for more efficient space complexity.