資料結構 PA1 Readme

姓名：馬健凱
學號：B07901108
系級：電機三

How to run my code

Type "***python3 pageRank.py***" will do!

Data Structure

I use lists for most of my data, and some are nested lists.
Test data:

1. ***test*** list of strings, saving **list.txt** commands

Page links:

1. ***pageToLinks*** list of list of int, saving the outbranching of the page
2. ***pageFromLinks*** list of list of int, saving all pages pointing to the page

ReverseIndex:

1. ***reverseIndex*** dictionary, where the strings map to their pages

PageRank:

1. ***prob_w*** list of int, ***PR*** in PageRank pseudocode
2. ***prob_r*** list of int, ***PRbefore*** in PageRank pseudocode
3. ***myPageRank*** dictionary, where the pages map to their rank

Top ten hit:

1. ***AND_set*** set, saving the AND result of the current query
2. ***OR_set*** set, saving the OR result of the current query

Algorithm

1. Save page links and ***reverseIndex***
2. Calculate PageRank (same as the pseudocode)
3. Output **pr_xx_xxx.txt**
4. Perform searches in **list.txt** (same as the search engine)
5. Output ***reverseIndex.txt***
6. Search engine
    a. Search
    b. Sort the results
    c. Output the top ten hits

Performance Analysis

Here we focus on calculating PageRank and the performance of the search engine. Denote the number of the pages as ***N***. We assume getting the count of the outbranching takes constant time.

Time complexity

PageRank: **O(N)**
Search engine:

1. Amortized **O(1)** for a search
2. Sorting takes **O(r lg r)**, where ***r*** is the length of the results in the sets.

Reference: What is the time complexity of checking if a key is in a dictionary in Python?

Space complexity

We store the links of the pages in nested lists, so the space complexity is about **O(N^2)**