Homework 2

1 Introduction

Now that we are more comfortable with reading from and writing to the command line, it's time to put those skills to work! ***Make sure you read the entire document***, there are some things even at the very end that you will not want to miss for this and future assignments!

2 The PageRank Algorithm

PageRank is an algorithm used by Google Search to measure the importance of different webpages. Named for one of the founders of Google, Larry Page, PageRank is still used to rank websites in Google's search engine results. In fact, the PageRank algorithm is what set Google apart from other popular search engines at the time of its introduction and is one of the main causes for its initial success. You can read even more about the PageRank algorithm here on Wikipedia.

The algorithm for PageRank is as follows:

$PR(A) = \frac{(1-d)}{n} + d(\frac{PR(T_1)}{L(T_1)} + \frac{PR(T_2)}{L(T_2)} + ... + \frac{PR(T_n)}{L(T_n)})$

where

- $PR(A)$ is the PageRank of an arbitrary page, A
- d is a damping factor which can be set between 0 and 1 (inclusive). It represents the probability that someone will click links from the page that they are currently on. The original damping factor used by Google was 0.85.
- n is the number of outbound links on page A
- $T_1 ... T_n$ are the pages linked to by Page A
- $PR(T_n)$ is the PageRank of T_n
- $L(T_n)$ is the number of outbound links on page T_n

3 Solution Description

You will be creating one file, PageRank.java, that will take user input to calculate a new page's PageRank.

You will need to ask the user for the following items:

- name of the page you are ranking

- number of outbound links it has
- which pages the user's entered page links to
- the damping factor

We are providing initial data on 10 different websites that you should account for in your code. These pages are shown in the table below.

Page Name	Number of Outbound Links	Predetermined PageRank
Wikipedia	10	4
Facebook	8	4
TechCrunch	6	3
Twitter	6	4
BlueApron	4	1
Instagram	1	1
Pinterest	7	4
Quora	5	3
GrubHub	2	2
Airbnb	4	2

For example, if a user says their page links to Wikipedia, you should use the data presented above to aid in the calculation. Each page has a name, a number of outbound links, and a predetermined PageRank, respectively.

When displaying the final PageRank to the user, be sure to include the page name they entered and its PageRank rounded to *two* decimal places.

Below is an example of what your user's interaction with the console should be:

```
Enter a page name:
stackoverflow

How many outbound links does stackoverflow have?
3

Available Pages:
Wikipedia
Facebook
TechCrunch
Twitter
BlueApron
Instagram
Pinterest
Quora
GrubHub
Airbnb

Which of the above pages does stackoverflow link to?
Wikipedia, Twitter, Quora

What damping factor would you like to use?
0.85

The PageRank of stackoverflow is: 1.47!
```

4 "Consider the following..."

Here are some tips to help you through this homework:

- **The API is your friend!** - You might find it beneficial to explore the String and Scanner classes to help you complete the assignment.
- If you find yourself bogged down by the PageRank equation, try to do the provided example on paper to figure out how to get to the number that was calculated based on the given inputs.
- Remember to put the collaboration policy at the top of `PageRank.java`. If you don't know what i'm talking about, please consult the syllabus.
- You do *not* have to worry about incorrect input. When running your code you can assume someone would give perfectly correct input every time.

5 Turn-in Procedure

Submit your `PageRank.java` file on T-Square as an attachment. Do not submit any compiled bytecode (`.class` files), the Checkstyle jar file, or the `cs1331-checkstyle.xml` file. When you're ready, double-check that you have submitted and not just saved a draft.

6 Verify the Success of Your Submission to T-Square

Practice safe submission! Verify that your HW files were truly submitted correctly, the upload was successful, and that the files compile and run. It is solely your responsibility to turn in your homework and practice this safe submission safeguard.

1. After uploading the files to T-Square you should receive an email from T-Square listing the names of the files that were uploaded and received. If you do not get the confirmation email almost immediately, something is wrong with your HW submission and/or your email. Even receiving the email does not guarantee that you turned in exactly what you intended.
2. After submitting the files to T-Square, return to the Assignment menu option and this homework. It should show the submitted files.
3. Download copies of your submitted files from the T-Square Assignment page placing them in a new folder.
4. Recompile and test those exact files.
5. This helps guard against a few things.
   (a) It helps insure that you turn in the correct files.
   (b) It helps you realize if you omit a file or files.[1] (If you do discover that you omitted a file, submit all of your files again, not just the missing one.)

[1] Missing files will not be given any credit, and non-compiling homework solutions will receive few to zero points. Also recall that late homework will not be accepted regardless of excuse. Treat the due date with respect. The real due date is 8PM Thursday. Do not wait until the last minute!

(c) Helps find last minute causes of files not compiling and/or running.