**P434 Project 1 - Pagerank**

Group 6 (Zhexu Fan & Chao Duan)

**Introduction of the main steps and data flow of the program**

In *public static void main(String[] args)* , it was calling four functions. These are exactly the main steps of the program:

       sequentialPR.parseArgs(args);

       sequentialPR.loadInput();

       sequentialPR.calculatePageRank();

       sequentialPR.printValues();

1. In *public void parseArgs(String[] args)*, we first get the input file name, output file name, number of iterations and the damping factor from the command line and save it to the project, so that we can use them for latter functions.
2. In *public void loadInput()*, we use the input file name from step 1 to open that file and load the content of the file into the program. In this step we basically filled up the hash map adjMatrix so the program can understand the input data.
3. *public void calculatePageRank()* was of course the main part of the project, it calculates the page rank of the project by using 2 main loops, and save the keys and result pagerank to another hash map called rankValues. The outer loop keeps track of how many iterations we need, and the inner loop was calculating after each iteration, what was the pagerank suppose to be.
4. *public void printValues()* actually contains of two tasks, first it sort the output hash map and save it to an array list, then it print the sorted result both to the output file and to the console.

**Sample output**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* The assignment#1 in P434 class \*

\* PageRank input Data \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

[chaoduan\_HadoopPageRank\_output.txt]

[11num of urls]

[6 of groups(Zhexu FAN & Chao Duan)]

e.g.: java PageRank pagerank.input iterations = 10 df = 0.85

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Top 10 is : \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

[2=0.3097295723325944, 1=0.3036191818465732, 4=0.06821469112858616, 5=0.03297885857776607, 3=0.03297885857776607, 0=0.027646509310404344, 10=0.01363636363636364, 9=0.01363636363636364, 8=0.01363636363636364, 7=0.01363636363636364]

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*all o sorted url:\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

[2=0.3097295723325944, 1=0.3036191818465732, 4=0.06821469112858616, 5=0.03297885857776607, 3=0.03297885857776607, 0=0.027646509310404344, 10=0.01363636363636364, 9=0.01363636363636364, 8=0.01363636363636364, 7=0.01363636363636364, 6=0.01363636363636364]