Assignment 7: MapReduce-based PageRank Algorithm

**by Vishal Doshi (vdoshi3@uic.edu)**

*The PageRankInputGenerator.java will generate a randon input file for you. Please provide number of pages.*

*NOTE: Make sure there is a directory named "pageRankInput" is created.*

*Also create directory named “unprocessedPageRankInput” input file: "pageRankInput.txt" [ie. unprocessedPageRankInput/pageRankInput.txt]*

**Steps:**

1. Extract and Import “PageRankHadoop.zip” into Eclipse.
2. Export as jar to desktop.
3. Start hadoop cluster and create “unprocessedPageRankInput” and “pageRankInput” directory in HDFS

./bin/hadoop fs -mkdir pageRankInput

./bin/hadoop fs -mkdir unprocessedPageRankInput

1. Put the “pageRankInput.txt” file ont desktop and then into “unprocessedPageRankInput” on HDFS

./bin/hadoop fs -put ~/Desktop/pageRankInput.txt unprocessedPageRankInput

1. Run the Jar. Provide "unprocessedPageRankInput/pageRankInput.txt " and “beta value (0.8 to 0.9)” as an argument while running it

./bin/hadoop jar ~/Desktop/jartest/pagerankhadoop.jar PageRankHadoop unprocessedPageRankInput/pageRankInput.txt 0.8

*Format: java <CLASS-NAME> <Input File location> <Beta Value>*

***I have also provided sample input, intermediate files generated and output files with the submission in respective folders.***

**Implementation Details:**

**PageRankHadoop.java**: The driver class. Preprocesses the input for dead end. Iterates until the ranks don’t fluctuates more than 0.05

**PageRankMapper.java**: Mapper class

**PageRankReducer.java**: Reducer Class. Calculates new ranks