Assignment 3

The zip contains

1)Source code file: PageRank.java

This file contains the source code for the page rank implementation

Four arguments need to be set. One for the input and remaining three for the different outputs of the map reduce jobs

The program has four map reduce jobs

The first job counts the total number of nodes in the given input file by counting the occurrence of data between <title></title> tags and calculating sum of this in the reducer of job 1.

The next map reduce job is to calculate the initial page rank of each page using 1/N and adding it to the value

The key will be the page name here and value will be the initial page rank appended by the list of outlinks for that page

The third map reduce job calculates the actual page rank. This job runs for ten iterations to get more accurate page rank.

The last map reduce job is required to sort the pages according to their page ranks in a descending manner.

2)SimplewikiOutput.txt

Contains top 100 output lines for Simplewiki.txt input file

3)WikiMicroOutput.txt

Contains top 100 output lines for WikiMicro.txt input file