Jimmy Tran

CSC 4760

21 March 2018

Assignment 4 Report

1) The lambda functions essentially create intermediate results by applying the steps of the PageRank algorithm without the need for many variable declarations. Step-by-step:

-The first function splits the given text file by whitespace into columns.

-The second assigns the first column of the file as the key and the rest as values to those keys.

-The third initializes the PageRank values for each key.

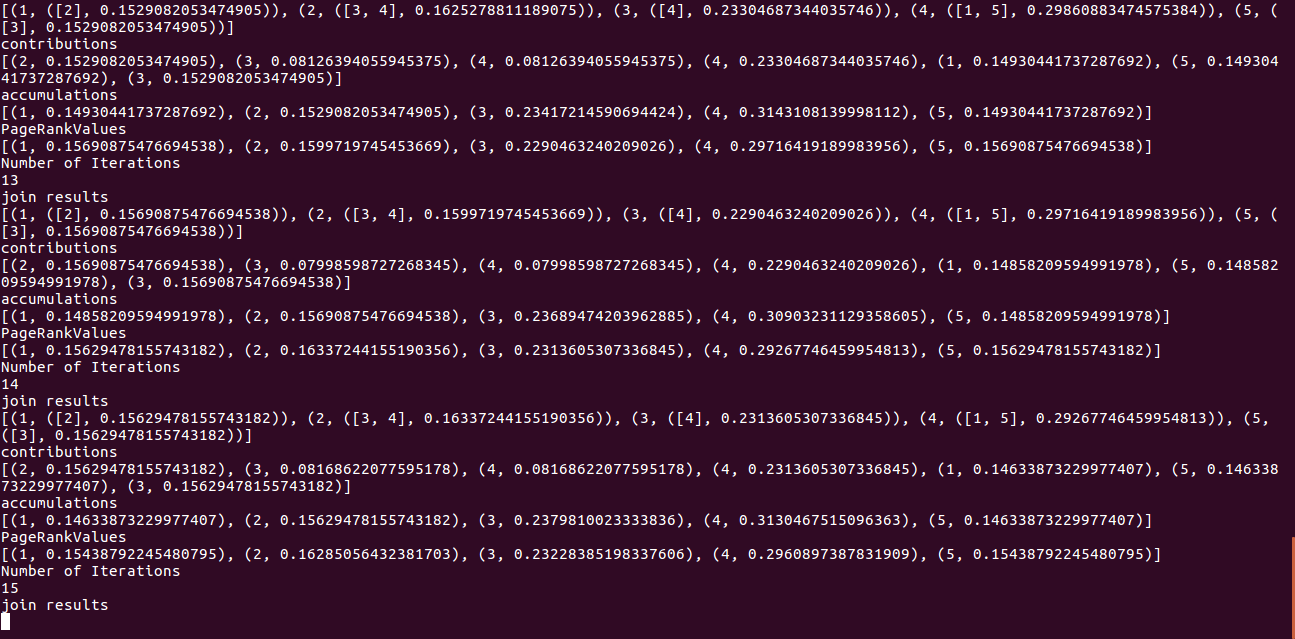
-The fourth creates a FlatMap RDD of the data.

-The fifth sums the x and y values to create a total value.

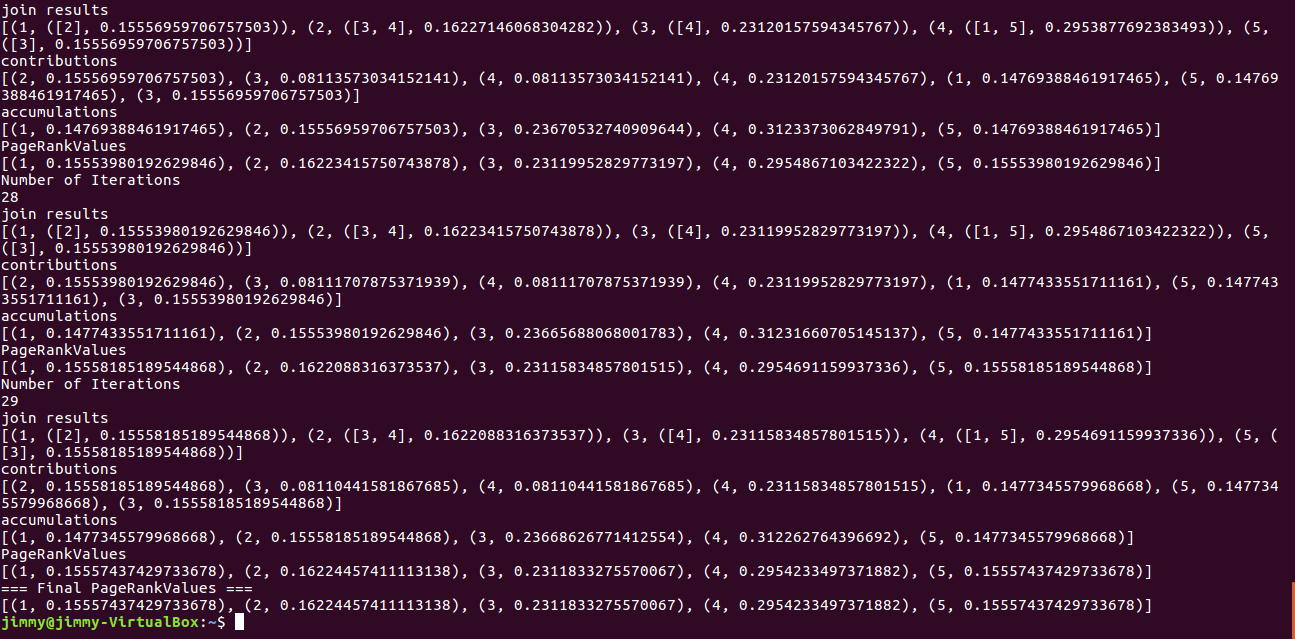
-The sixth applies the PageRank formula with c=0.85 and gives the final results.

Most of these intermediate results are (key, value) pairs.

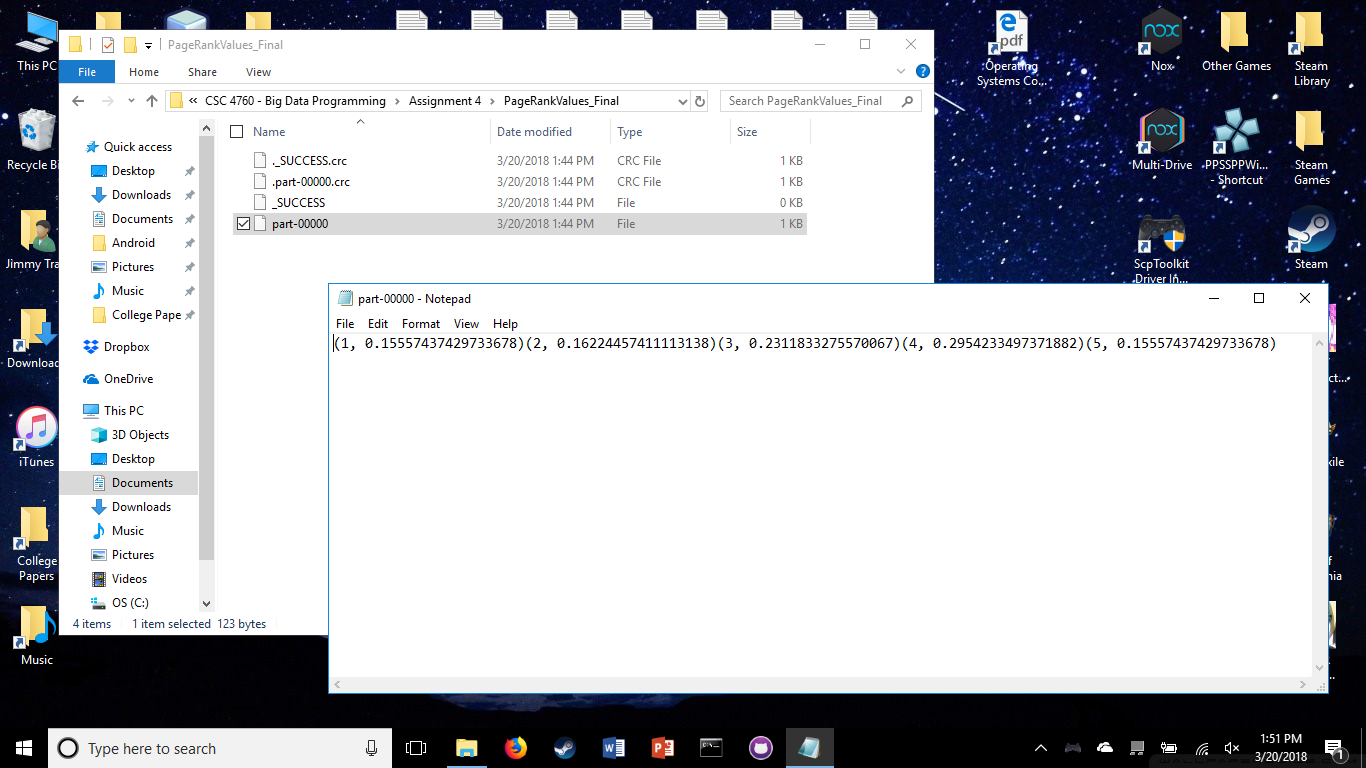
2) Screenshots of results (Intermediate – This was around the 15th iteration)



Final Outputs:



Output file on local filesystem (It’s in my normal Windows OS but that doesn’t really matter. I just moved the file from the VM Linux):



Result Report:

This implementation of PageRank gives the exact same values as given in the Assignment 4 text file which means I completed the code correctly.