Problem: To compute PageRank for the given species dataset.

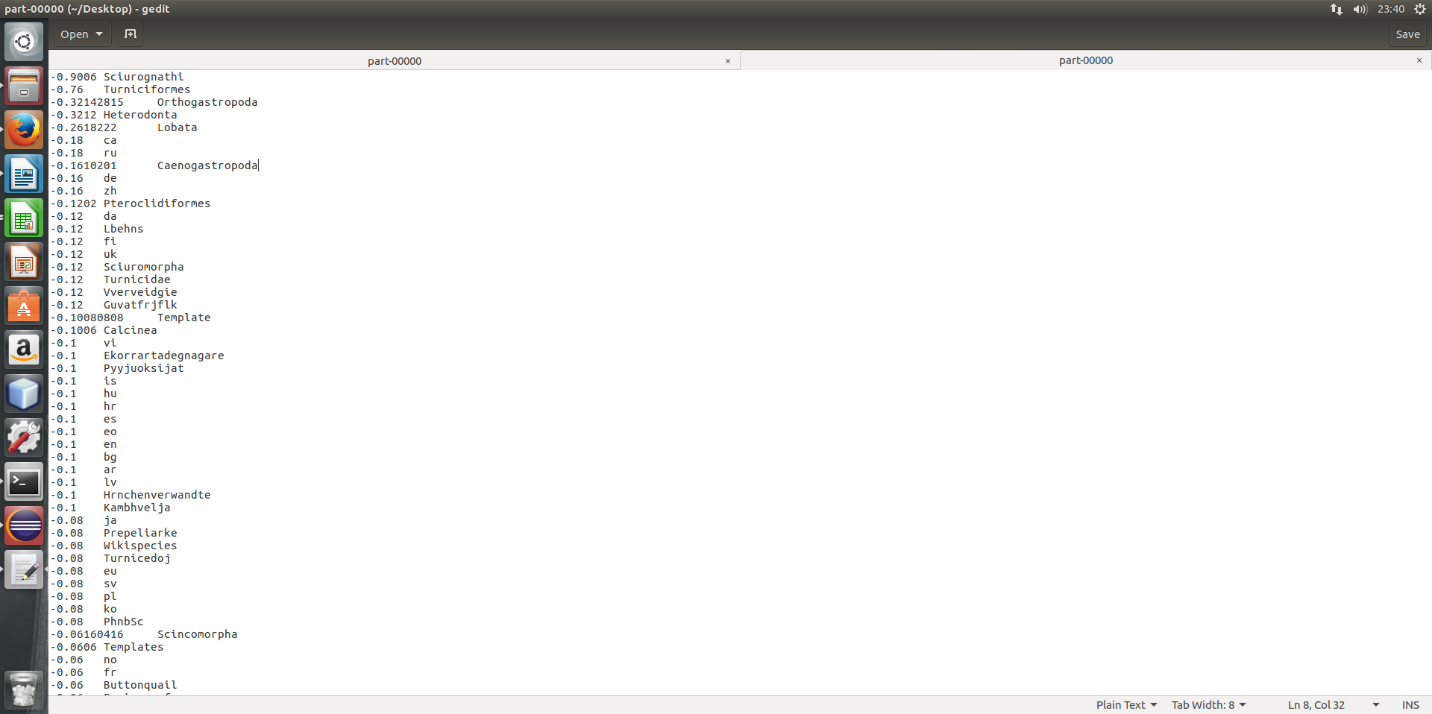
Steps followed:

1. Reduced the driver to single driver class.
2. Created a while loop to perform iteration.

To execute the code:

1. Open the file in eclipse mars.
2. Set Run configuration input as : <input> <inter\_output\_1> <inter\_output\_2> <output>
3. Run the job.

I have attached the output file herewith.



The input is converging for a damping factor of 0.01 and iterations 10.

Lesser the damping factor, higher is the rate of convergence.

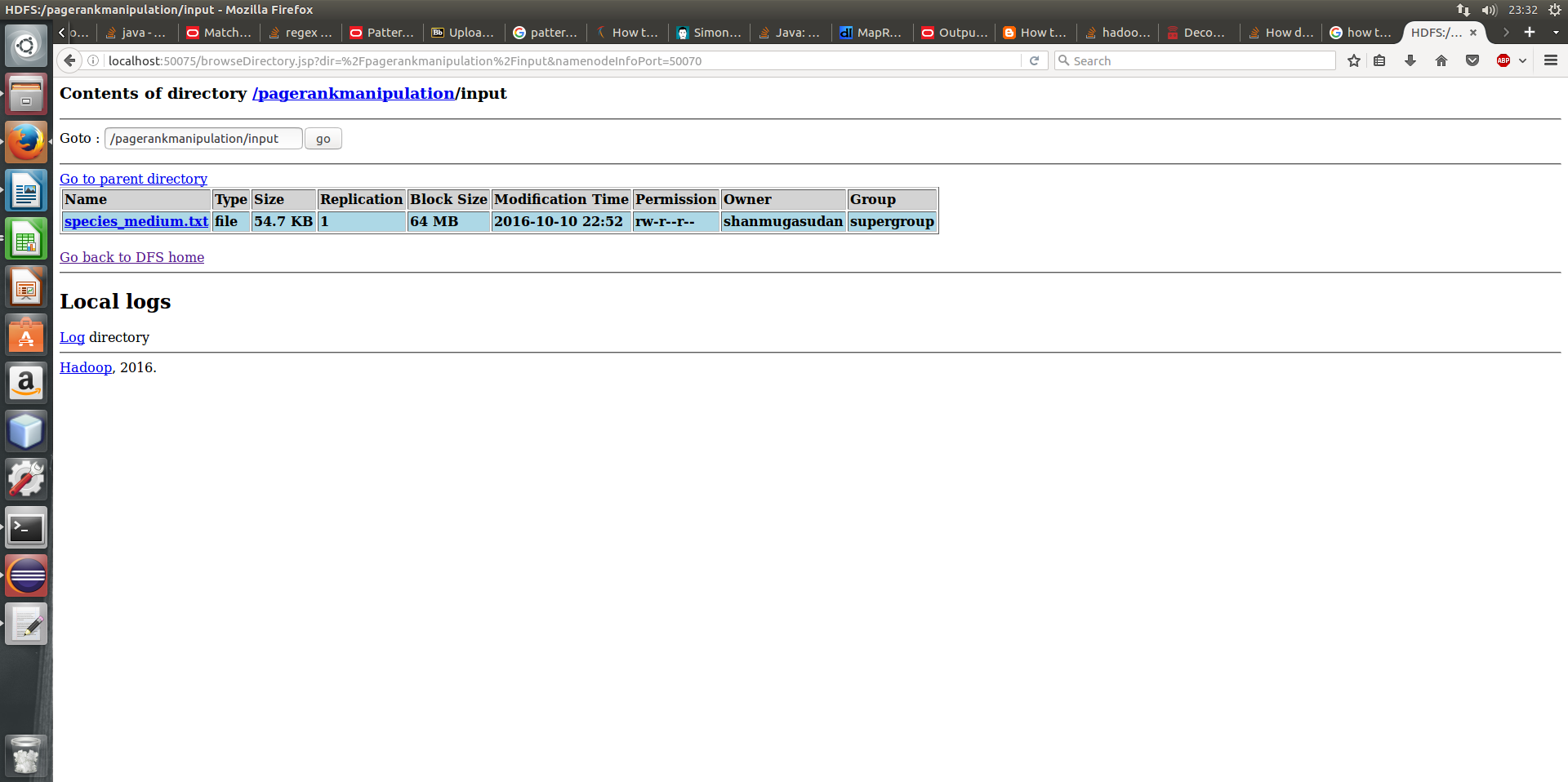
However, after 25 iterations, I see that the values diverge.

My results contained empty values, number values and non-alphabetic (English) characters.

I have used Regex pattern, trim function and isempty function to check and replace, eliminate the junk records. The generated output is free of any junk data.

Pseudo Distribution Mode:

Loaded the file,



I have managed to load the file to hdfs. However, I faced execution errors.

