**Instruction for compiling and running Page Rank Algorithm**

* Files: Zip file comes with following files
  1. *PageRank.py:* python file implemented on python 3, which is implementation of page rank algorithm.
  2. *wt2g\_inlinks*: inlinks file for WT2G collection as given in problem set.
  3. page\_rank\_ABCDEF: output required for assignment.
  4. Analysis of Page Rank results: an analysis of the PageRank results obtained
* Steps to run PageRank.py

>python3 PageRank.py “wt2g\_inlinks”

Run Time: ~100 seconds

* Pre-requisite: python version 3.x (preferable 3.5.0)
* Output: (PageRank.py generates following output files)

1. perplexity\_values: a list of the perplexity values obtained in each round until convergence
2. unsorted\_page\_rank\_values: pages with their page rank values in wt2g collection
3. sorted\_page\_rank\_values: pages sorted in descending order based on their page rank
4. pr\_results:
   1. a list of the document IDs of the top 50 pages as sorted by PageRank, together with their PageRank values;
   2. a list of the document IDs of the top 50 pages by in-link count, together with their in-link counts;
   3. the proportion of pages with no in-links (sources);
   4. the proportion of pages with no out-links (sinks); and
   5. the proportion of pages whose PageRank is less than their initial, uniform values.

* Other Outputs:

1. page\_rank\_ABCDEF: list of the PageRank values obtained for each of the six vertices graph after 1, 10, and 100 iterations of the PageRank algorithm.
2. Analysis of Page Rank results: an analysis of the PageRank results obtained

***References***:

* Python Tutorials: <http://www.tutorialspoint.com/python/>
* Page Rank algorithm: <https://www.youtube.com/watch?v=4c3DAxQXzLI>
* Python Docs: <https://docs.python.org/>

Note:

1. I have provided all the output files generated by PageRank.py in this zip file for your reference but in actual these output files are generated at runtime as per given inlink input.