Reverse Web-Link Graph

Map Reduce Job:

1. Import the required packages, MRJob, MRStep, and re.
2. To skip the lines starting with "#", a pattern (r"^(?!#).*") is defined using re which considers only the lines without "#".
3. The mapper function creates a list of source and target web pages for each line of document (['source ID', 'target ID']) and returns the reversed list (['target ID', 'source ID']).
4. The reducer function returns a list of sources for each target, i.e., for each target, the reducer generates a list of the web pages (the sources) to which the target is linked.
5. The mapper and the reducer are combined via MRStep (the steps function).

Commands to be executed in terminal:

1. The results are stored in a text file called 'reversed_web_link.txt' using the below command:

   python reverse_web_link/map_reduce_web.py reverse_web_link/web-Google.txt > reverse_web_link/reversed_web_link.txt

To open the "reversed_web_link.txt" file, please copy it to your operating system environment and open from there, not in your IDE like PyCharm