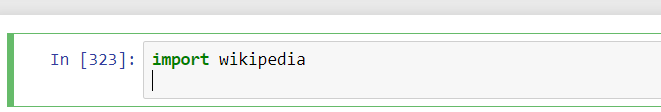
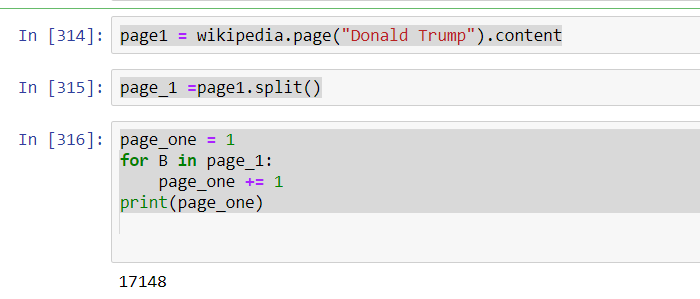
**COMMENTS ON CODE:**

Two URLS from source data is come from:

* <https://en.wikipedia.org/wiki/Donald_Trump>
* <https://en.wikipedia.org/wiki/Panama_Papers>



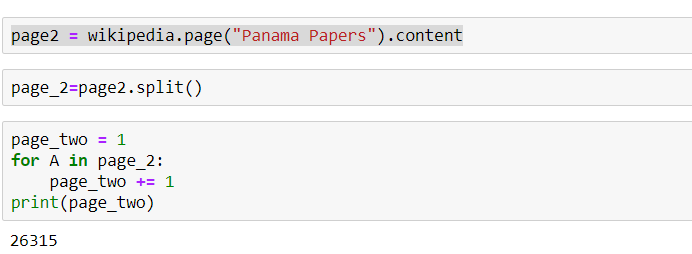
1. Installing wilipedia package.



2.Make a variable to load Wikipedia page content(Donald Trump).

Split the page word by word.

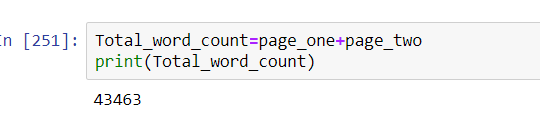
For loop counts all words with the increment of each word till the end.

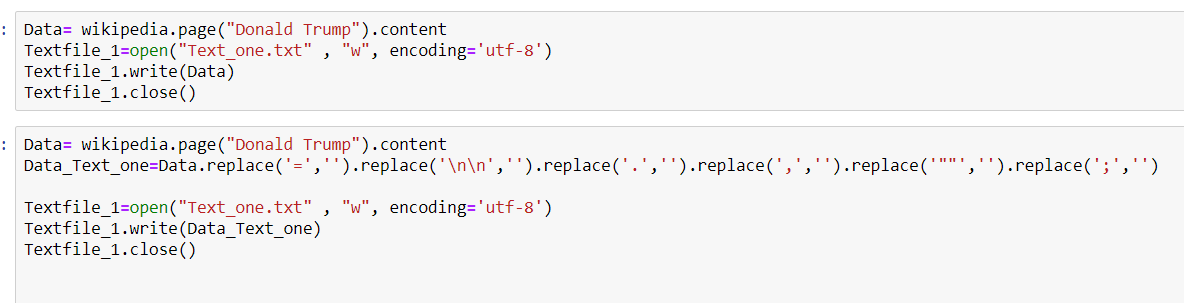


3.Make a variable to load Wikipedia page content(Panama Papers).

Split the page word by word

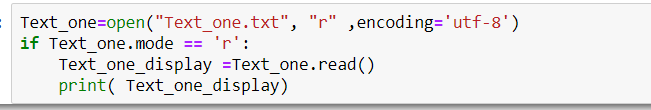
For loop counts all words with the increment of each word till the end.



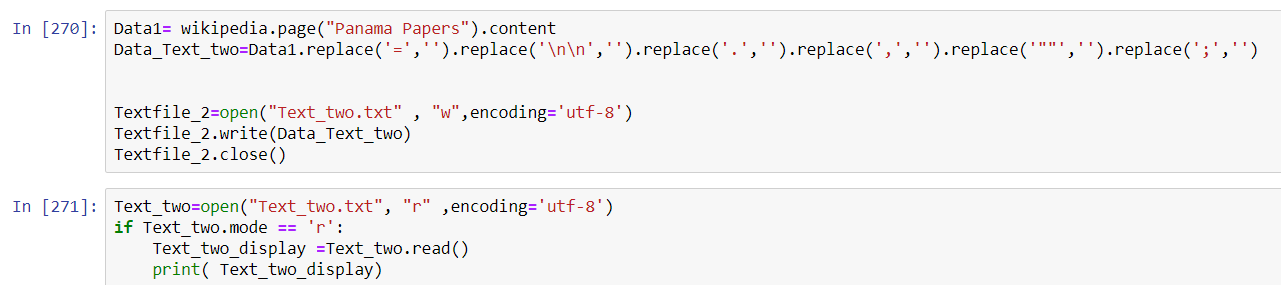


4. Created a new file with extension of (Text\_one.txt), write the data of page “DonaldTrump” on it. Write the data of page in to that text file.

Clean the text file with extra spaces, commas, full stops and =. And write the clean data in to “Text\_one.txt”.



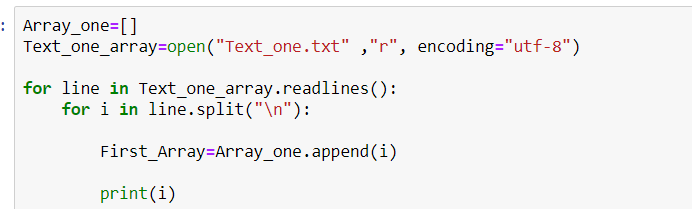
5. Text\_one.txt file was on write mode ,Change it in to read mode to see the display of data.



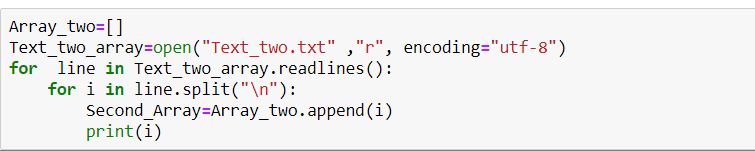
6. Created a new file with extension of (Text\_two.txt), write the data of page “Pananma Papers” on it. Write the data of page in to that text file.

Clean the text file with extra spaces, commas, full stops and =. And write the clean data in to “Text\_two.txt”.

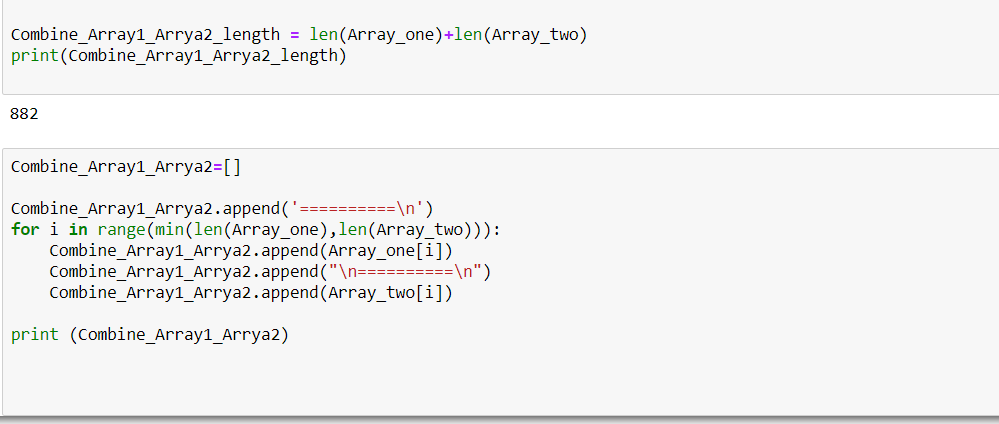
Text\_two.txt file was on write mode ,Change it in to read mode to see the display of data.



7. Load the Text\_one.txt data in to “ARRAY”(Array\_one).Every time when there is new line array is going to append and go to the next position.



8. Load the Text\_two.txt data in to “ARRAY”(Array\_two).Every time when there is new line array is going to append and go to the next position.

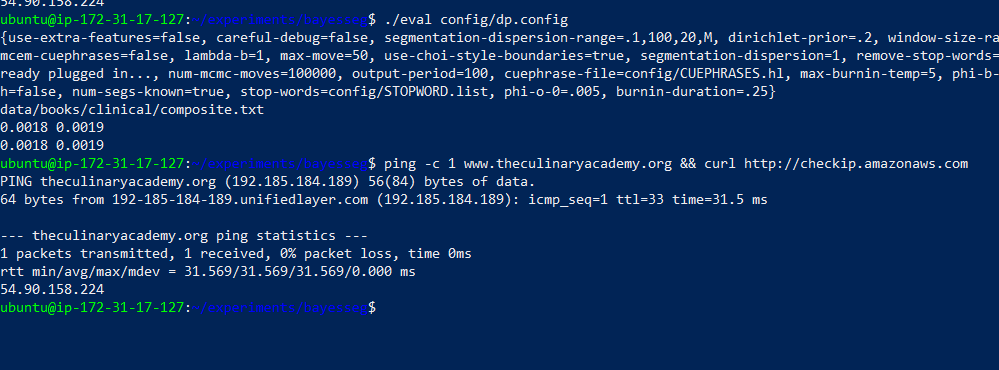


9. Generated a third array named “Combine\_Array1\_Array2”to load Array\_one and Array\_two in sequence wise with the append of “==========”.First Array\_one segment 1 will come (after the top line of “==========”) than again “==========” will come and next line Array\_two segement 1 will come.This pattern is keep going till the any one of array minumum length will not kick in.



10. Generating a new text file (composite.txt) for Array three which is [combine\_Array1\_Array2] ,write the data of Array three in to that text file (composite.txt).Read and display it to see whether the combined data set looks perfect to load it in to bayesseg segmentation.

**SCORE OF COMPOSITE.TXT**

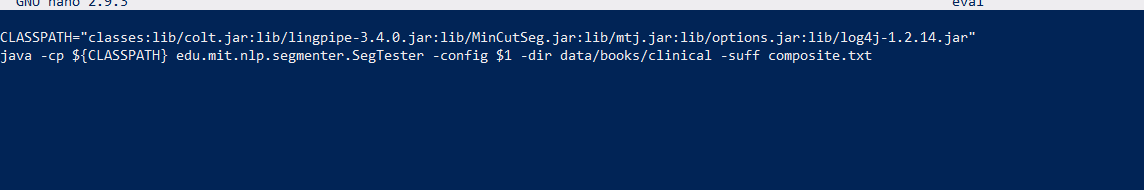


Scores produced by the program of composite.txt look significant as these are 0.0018 and 0.0019(as approximately near to ‘0’).The data has taken from two different pages from Wikipedia (topic wise).First one “Donald Trump” and Second one “Panama Papers”.

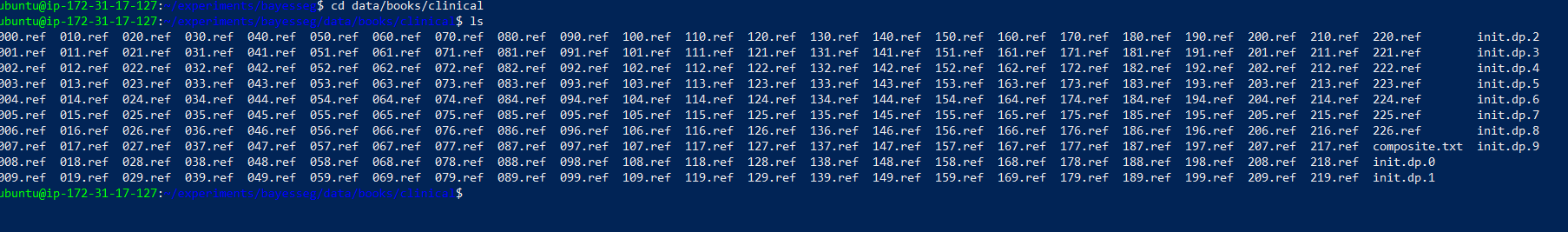
All segmentations are quite different from each other.They are systematically assembled each other document and segment wise.

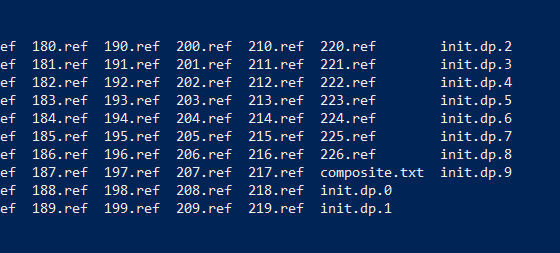
doc1\_seg1 doc2\_seg1 doc1\_seg2 doc2\_seg2 ... doc1\_segN doc2\_segN

**CHANGING EVAL FILE DIRECTORY FROM –SUFF 1.REF TO composite.txt**

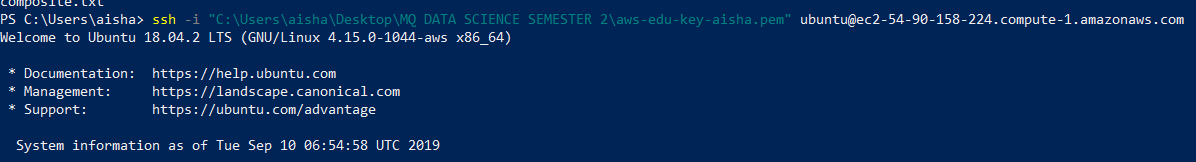


**ALL FILES UNDER CLINICAL DIRECTORY**

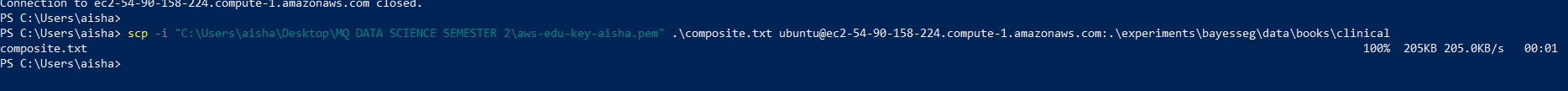




**SSH COMMAND IS CREATING CONNECTION BETWEEN VIRTUAL MACHINE AND LOCAL (POWER SHELL)**



**SCP COMMAND IS TRANSFERRING COPY FROM LOCAL STATION TO EC2.**





**READING DATA INSIDE THE composite.txt**

