1. Install ubuntu
2. Install eclipse
3. Import wordcount project
4. Add hadoop jar as external libraries (Right click on project -->properties--->java build path--->add jar
5. For wordcount.java

1)Right click on file → run as coniguration

2) click on argument tab

3) pass two argument (path for input file/folder , path for output folder

6)change the txt file obtain in above step in CSV ( using code in nydatawebcrawler.ipynb)

7)repeat above for all Data

8)Copy the csv files created above in wordCloud-demo folder

9)change the filename in js file accordingly

10) open index.htm in browser

11)steps to find top ten co-occurence word

1. Run .java file
2. Input --- > (file path of input folder/file ,path to output folder )
3. Convert the o/p of above step in to csv
4. Run .java file
5. Input ---> (file path created in step C , path to o/p file.csv)

12) copy above csv to Top10cloud.demo .Change filename accordingly

13) open index.htm in browser