README file

geoLondonJan23.json file contain geo data for London

the fields of interest are:

text and coordinates

Directory

credModelFiles – this is for task 2

this directory contains three files

1.highQuality.json

2. lowQuality.json

These files contain data in

{"\_id":{"$oid":"623cb1491fdfc2cfea3b85c4"},"username":"miguel\_zavalaa","description":"•Mind with no sense in it• •Stay away from me ‘cause I’m dancing to quite a different drumbeat•","qualityS":0.6081996434937612,"text":"Turisteando (@ Piccadilly Circus in London, Greater London) https://t.co/ear25pmZNc https://t.co/528VOdxIeF"}

“\_id” is unique tweet id

“text” is tweet text

“qualityS” is the credibility score computed by the algorithm we discussed in the class

For the task ii; we need only the text data (you are free to exploit credibility score if you wish)

Process these text data and build a dictionary

Third file contains the background data - For legacy reasons this is in a different format

{"\_id":"1613006911946776576","text":["confidence","dont","fix","lost","need","negative","playing","right","support","well"],"score":0.6367300158372926}

“\_id” is unique tweet id

“text” is tweet text

“score” is the credibility score computed by the algorithm we discussed in the class

You can process the text field by reading into a dictionary