COMP6930 - Assignment #3
(Due Date: 6$^{\text{th}}$ April 2018 @ 11:59 PM)

Instructions: You are to submit a **.zip** folder named using your UWI ID number to inzamam.rahaman@outlook.com.

Part A - Clustering Applied for Topic Mining (30 marks)

The file *tweets.json* contains 251 (not necessarily unique) tweets concerning the 2013 Boston Marathon Bombing. Using a clustering method, it should be possible to group related tweets together. Using an appropriate clustering method, partition the tweets in the tweets.json file into mutually exclusive clusters. (Hint: It may be useful to think of each tweet as a bag of words where word order is irrelevant to meaning).

Moreover, after splitting the the tweets into clusters, write for the five largest clusters a brief description of the topics/ideas/concepts which can describe the clusters. Depending on the clustering method you use, you may compute noise clusters. You can ignore said noise clusters.

Note that a Python file, *parser.py*, has been provided that reads *tweets.json* into a Python list.