6. Clustering Tweets

During this lab, you are going to cluster the tweets of the previously given dataset[1] to create groups of similar tweets. This task should be completely unsupervised. The recommended algorithm for this task is k-means. Concretely, the steps to follow are:

1. Implement k-means
2. Cluster the tweets using k-means
3. Compare the resulting clusters with the classes defined by the hashtags.

Deliverable: Write a report of 1-2 page that consists of (1) a very brief description of your implementation of K-means (2) the different clusters (3) A comparison between the clusters and the original classes (hashtags)
Deadline: 22/01/2016 at 23:55 am.

[1] http://www.inf.unibz.it/ mkacimi/teaching.shtml