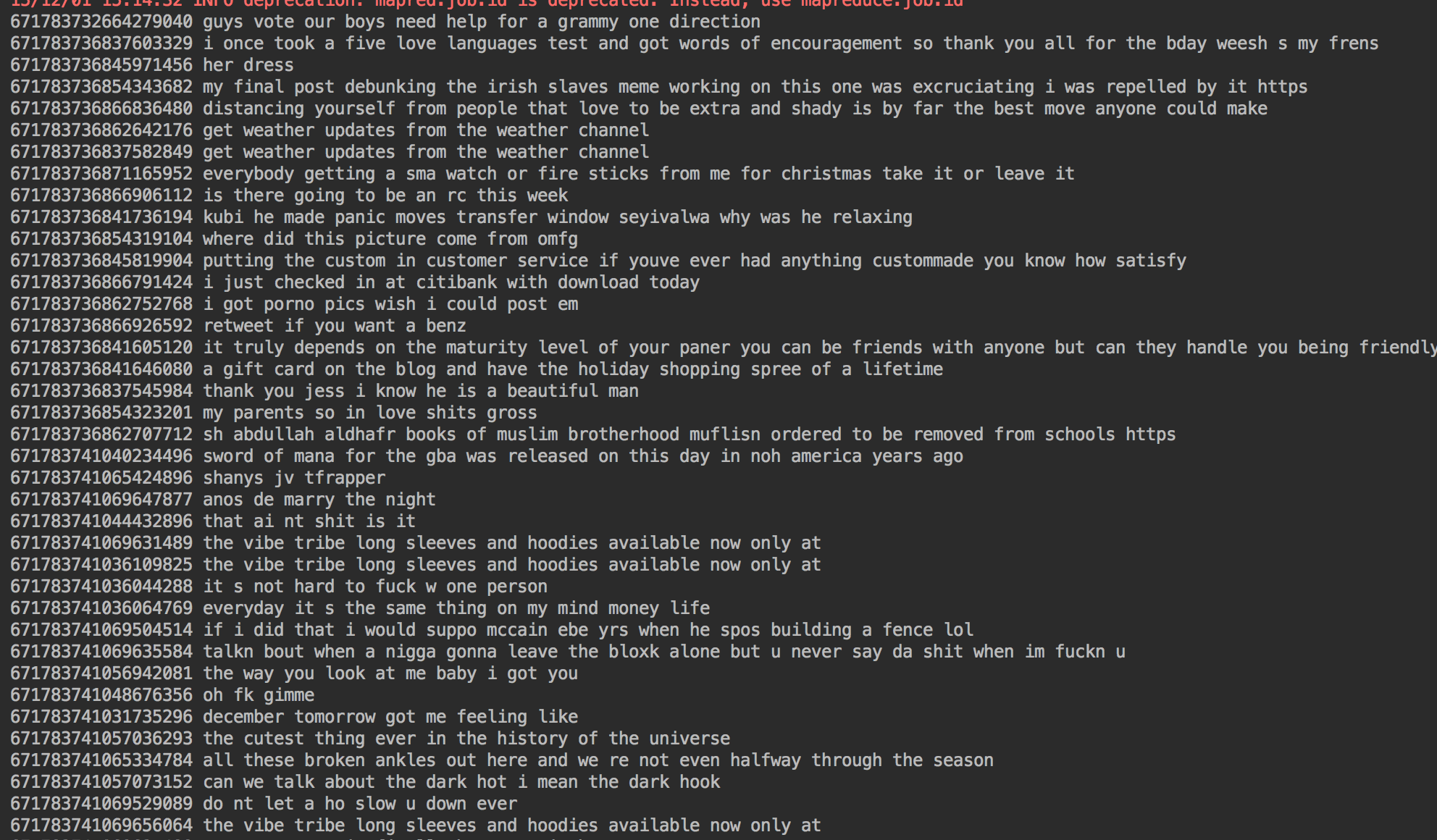
ScreeShots For Clustering Operation outputs for each step

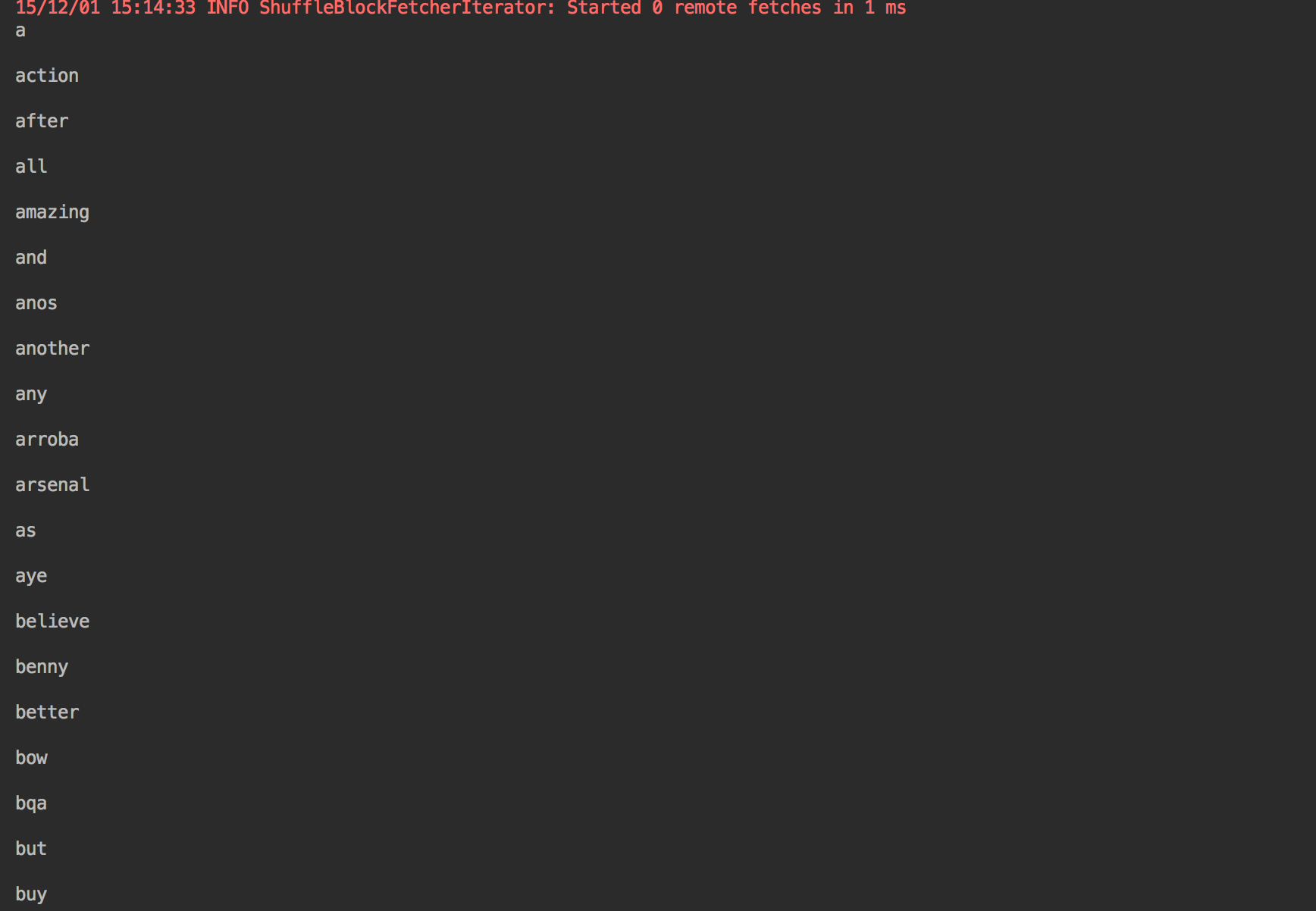
1. read File from the path

data format is String including the ID for each tweets and content after tokenization.

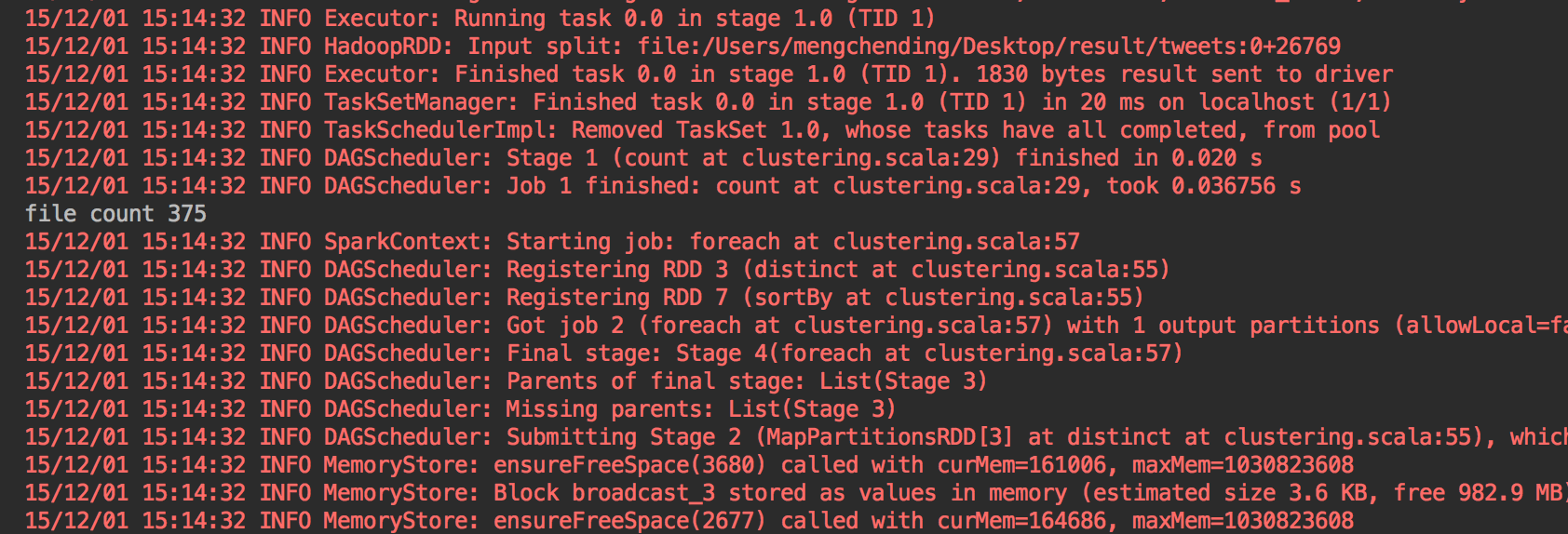


2. MapReduce the data

2.1 Get Dictionary

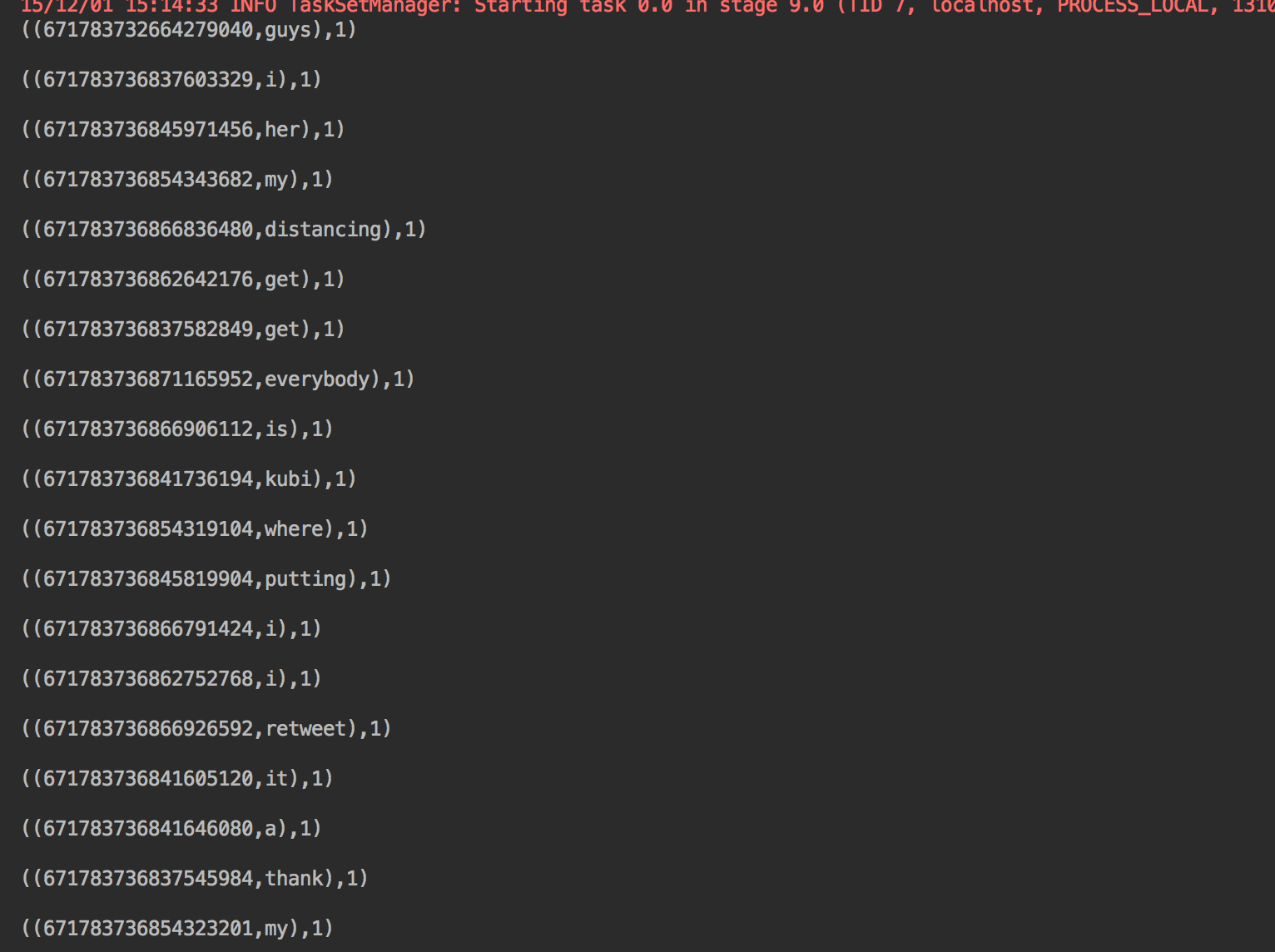


2.2 Calculate the number of documents in the dataset



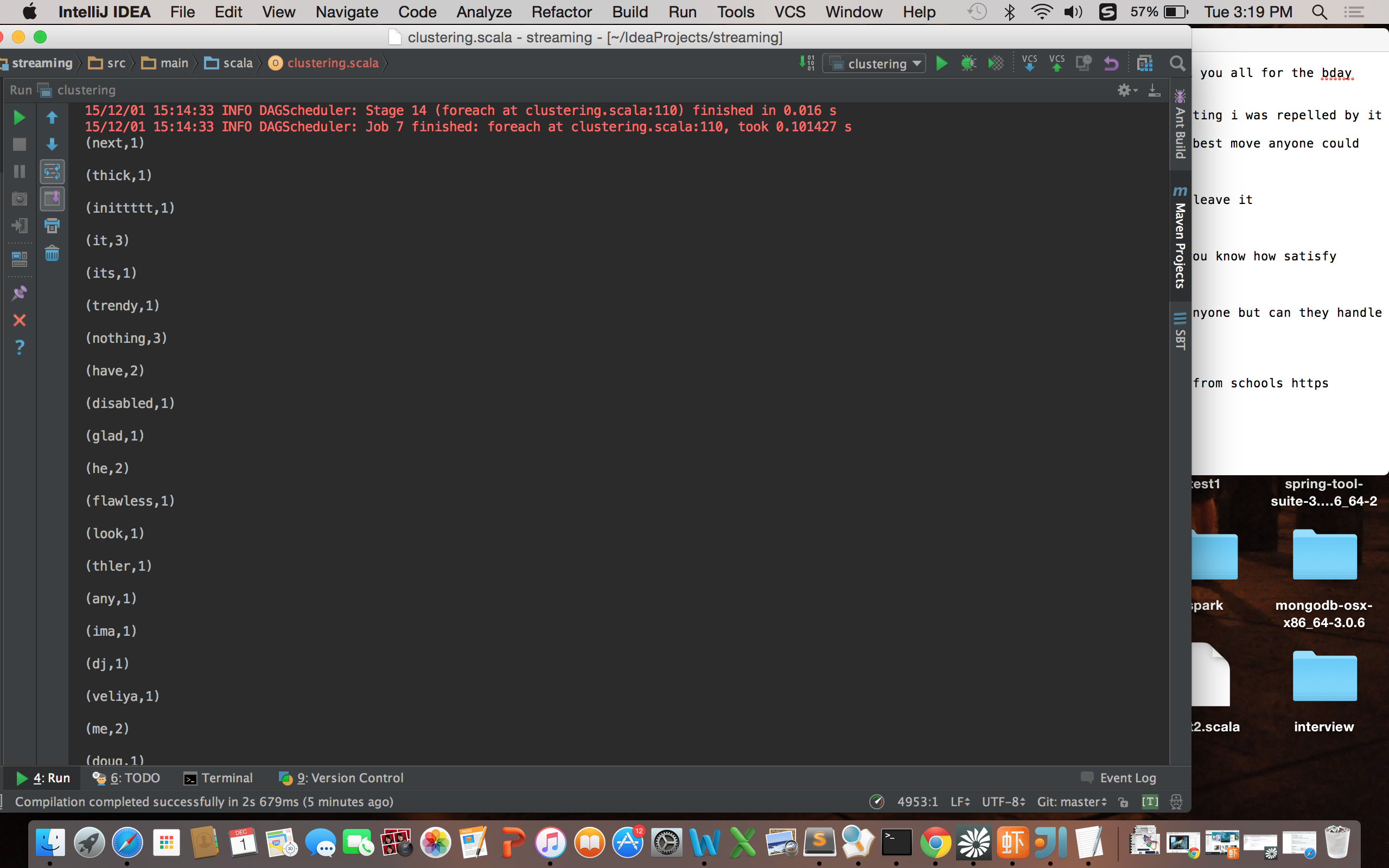
2.3 Calculate the numbers of each term occurrences in the document.

Data Format is ((id,term),frequencies)



2.4 Calculate the number of occurrences for each term in the dataset

Data Format(term,frequencies)



2.5 Use the K-means cost (sum of squared distances of points to their nearest center) for this model on the given dataset as the finishing signature.

