

Exponentia Data Labs Interview Task

Description :

Given file (as_tweets.txt) contains 100,000 tweets about the movie “**American Sniper**”. These are unprocessed real tweets. As you may know, the movie was directed by **Clint Eastwood**, starring **Bradley Cooper** and was based on the life story of **Chris Kyle**.

First task is to break the twitter hashtags into proper words, for example:

`#goldenglobes -> golden globes,`

`#siennamiller -> sienna miller`

... Example 1

Your second task is to find out what people have said about these three personalities (**Clint Eastwood, Bradley Cooper and Chris Kyle**), in other words find out the top trending topics about these people in the current dataset. A trend can be a single word (`stunning`) or may comprise of multiple words (`looks stunning`) as shown in Example2 below. Show 5 top trends per person. The tweet count should be in decreasing order (use example 2 format). Words with similar meaning should be considered same, i.e., `stunning`, `superb`, `wonderful` and `awesome` should be treated as one word; you can pick whichever word you like

`looked awesome(23tweets)`

`looked wonderful(7tweets)`

should be clubbed as `looked wonderful(30tweets)` or `looked awesome(30tweets)`

PS : Final top trends should be relevant and meaningful

For example, consider these 3 tweets about **Sienna Miller** from the same dataset:

1. Sienna Miller is stunning!!! #GoldenGlobes She brought the heat in @AmericanSniper!!!
2. @AmericanSniper and @FoxcatcherUK actress #SiennaMiller looks stunning in #Forevermark diamond jewelry. <http://t.co/2Y8JvJ8KlF>
3. Watching #AmericanSniper awesome acting by @BradleyCooperUS and @SiennaMillerr playing the life of Chris Kyle and Taya a ten

So for Sienna Miller one trending topic can be “looks stunning” or just “stunning” and another can be “awesome acting”, this can be represented as follows:

Sienna Miller

... Example 2

stunning (2tweets)

awesome acting (1tweet)

Remember these are not the only trends which can be extracted from the above 3 tweets. Other trends may include Golden Globes, diamond jewelry, Forevermark, FoxcatcherUK etc. Use your own judgement to figure out meaningful trends.