

Hash Apart

Nuvento

April 5, 2013

A hashtag is a word or a phrase prefixed with the symbol #, a form of metadata tag. Short messages on microblogging social networking services such as Twitter, Tout, identi.ca, Tumblr, Instagram, or Google+ may be tagged by including one or more with multiple words concatenated, e.g.: #Wikipedia is an #encyclopedia. Hashtags provide a means of grouping such messages, since one can search for the hashtag and get the set of messages that contain it.

In the problem you are given a collection of hashtags. In the data set you will see that many hashtags are concatenation of words or numbers.

For example: #ideaForIndia has three distinct words idea, For and India. Objective will be to break each hashtag as collection of meaningful words.

Your output file shall be of the form:

TAG : comma separated words, e.g.

#ideaForIndia : idea, For, India

Attach following when you reply with answer.

1. A document explaining solution approach
2. code
3. output as a text file in the format stated above
4. You are free to use any available code or refer and use anything in Internet or otherwise - however, please make a note of that in the solution approach document.

Best wishes!