Flickr BOG class

Bag of words is a representation for text analysis that counts each instance of a word. This program goes through all the tags for each photo and counts how many times they are encountered. It has two methods for data processing, get\_tag\_info and tags\_to\_BOG.

**get\_tag\_info**

Takes in csv and a ‘tag’ and outputs a row for the location of each tag. Only one tag can be pulled at time. To access the specific row/columns with pandas the csv must have the column configuration shown below. Column names can be altered but the column order must be preserved, or the code needs to be adjusted for the new column position. Output from FlickrResearch is already in the correct configuration.

Figure 1: Sample input csv file

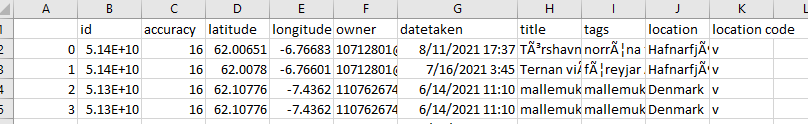
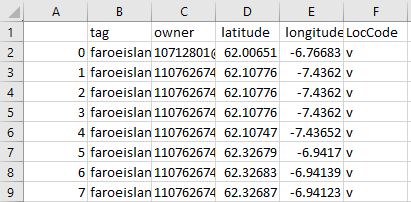


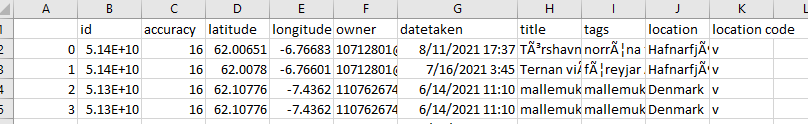
Figure 2: Outputs a row for each instance of a tag. This output is for mapping tags in a geographic information system. This output contains 87 tags see figure 4.



**tags\_to\_BOG**

Takes in a csv file. To access the specific row/columns with pandas the csv must have the column configuration shown below. Column names can be altered but the column order must be preserved, or the code needs to be adjusted for the new column position. Output from FlickrResearch is already in the correct configuration.

Figure 3: Sample input csv file



Outputs a Bag of words representation for tags and some additional information. Information provided includes tag, tag count, list of owners and a break down of locals and visitors. Nan = not a number and represents photos that have no tags. The sum of the local count and the visitor count should equal the owner count. The output is not sorted in any way, the results pictured below were sorted in excel.

Figure 4: Sample output

