

```
#!/usr/bin/env python3

# Author: Maharnav Singhal

import nltk

from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer
import os

stemmer=SnowballStemmer("english")
STOPWORDS=set(stopwords.words("english"))

def write_counts(counts,filepath):
    with open(filepath, "w") as outfile:
        for word in counts.keys():
            to_write=word+": "+str(counts[word])+"\n"
            outfile.write(to_write)

def get_counts(infilename):
    with open(infilename,"r", encoding="utf8", errors='ignore') as
infile:
        text=infile.read()

        tokens=word_tokenize(text)
        words=[word.lower() for word in tokens if word.isalpha()]
        words=[stemmer.stem(word) for word in words if word not in
STOPWORDS]

        counts={}
        for word in words:
            counts[word]=counts.get(word,0)+1

        sorted_tuples=sorted(counts.items(),key=lambda item: item[1],
reverse=True)
        sorted_counts={k:v for k, v in sorted_tuples}
        return counts

#replace the folowing path according to your work-environment

filenames=os.listdir("/Users/maharnavsinghal/Library/Mobile
Documents/com~apple~CloudDocs/Programming/HSL 1/Bag of Words/
Files/")
for filename in filenames:
    counts=get_counts("/Users/maharnavsinghal/Library/Mobile
Documents/com~apple~CloudDocs/Programming/HSL 1/Bag of Words/
Files/"+filename)
    write_counts(counts,"/Users/maharnavsinghal/Library/Mobile
Documents/com~apple~CloudDocs/Programming/HSL 1/Bag of Words/
Results/"+filename)
```