VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

Register Number: 18BCE0745

Name: Gourishetty Sreemanth

Code:

```
# -*- coding: utf-8 -*-
"""
Created on Thu Jul 23 16:13:45 2020

@author: Sreemanth
"""
import re
from nltk.corpus import stopwords

from nltk.tokenize import word_tokenize



with open('Artificaial intelligence.txt', 'r') as file:
    data = file.read().replace('\n', '')

stop_words = (stopwords.words('english'))
```

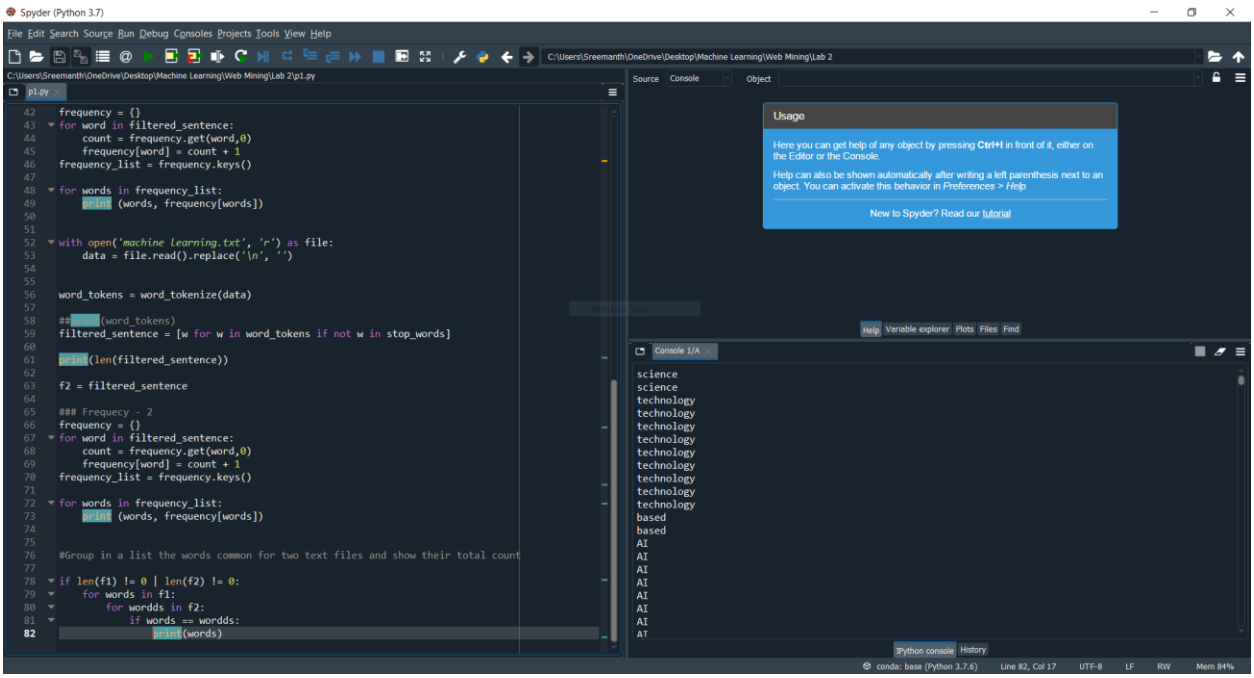
```
stop_words.append('a')
stop_words.append('they')
stop_words.append('the')
stop_words.append('his')
stop_words.append('.')
stop_words.append(',')
stop_words.append('so')
stop_words.append('and')
stop_words.append('were')
stop_words.append('from')
stop_words.append('that')
stop_words.append('of')
stop_words.append('in')
stop_words.append('only')
stop_words.append('with')
stop_words.append('to')
word_tokens = word_tokenize(data)


filtered_sentence = [w for w in word_tokens if not w in stop_words]


print(len(filtered_sentence))


f1 = filtered_sentence
### Frequecy - 1
frequency = {}
for word in filtered_sentence:
   count = frequency.get(word,0)
   frequency[word] = count + 1
```

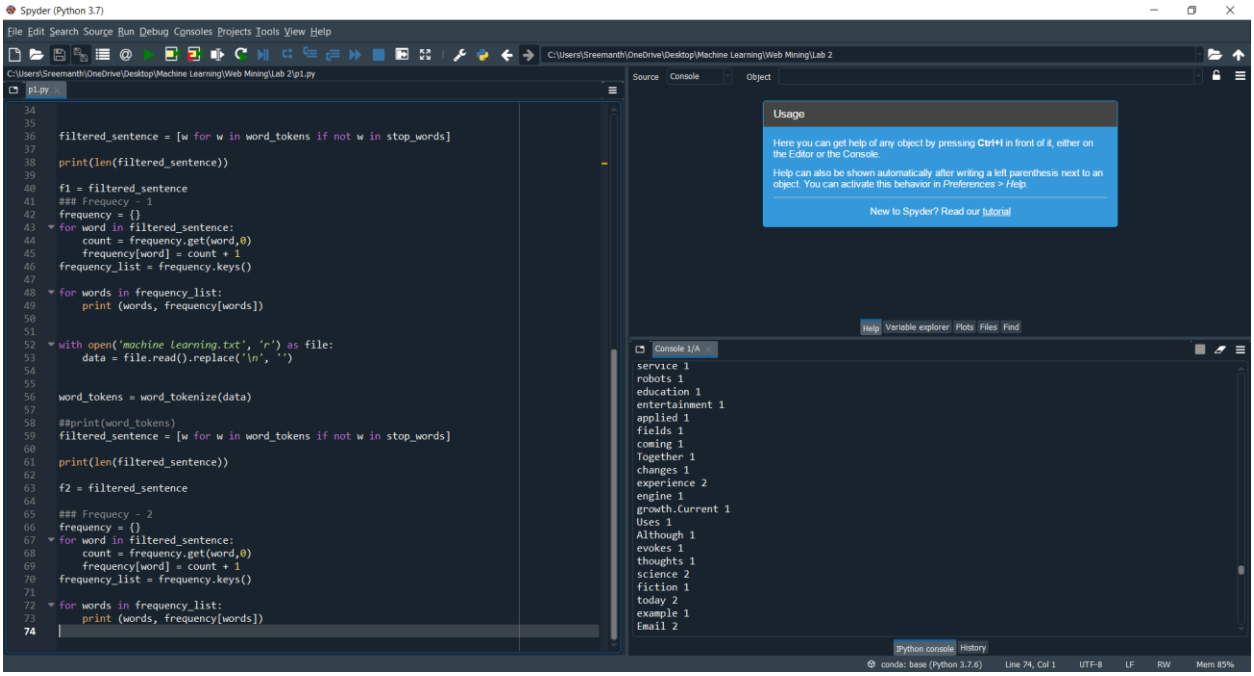
```
frequency_list = frequency.keys()

for words in frequency_list:
    print (words, frequency[words])



with open('machine learning.txt', 'r') as file:
    data = file.read().replace('\n', '')



word_tokens = word_tokenize(data)

##print(word_tokens)
filtered_sentence = [w for w in word_tokens if not w in stop_words]

print(len(filtered_sentence))

f2 = filtered_sentence

### Frequecy - 2
frequency = {}
for word in filtered_sentence:
    count = frequency.get(word,0)
    frequency[word] = count + 1
frequency_list = frequency.keys()

for words in frequency_list:
    print (words, frequency[words])
```

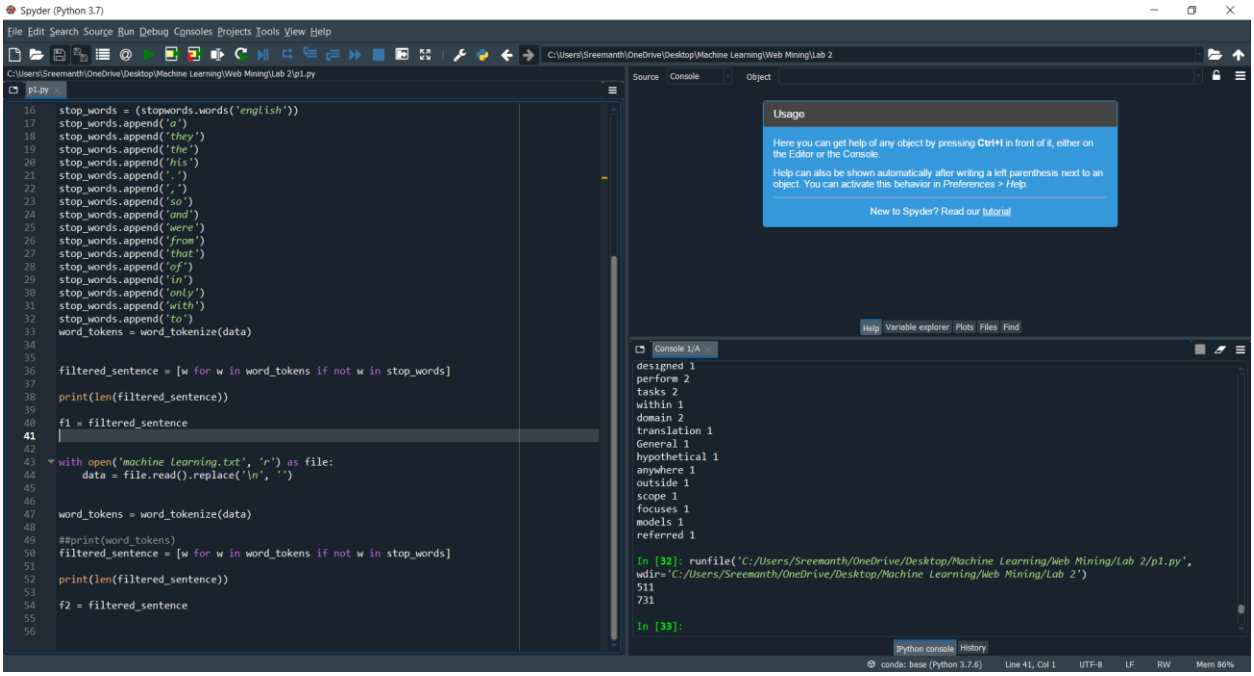
#Group in a list the words common for two text files and show their total count

```
if len(f1) != 0 | len(f2) != 0:
    for words in f1:
        for wordds in f2:
            if words == wordds:
                print(words)
```

Screenshots:




Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\Sreemanth\OneDrive\Desktop\Machine Learning\Web Mining\Lab 2\p1.py
p1.py
filtered_sentence = [w for w in word_tokens if not w in stop_words]
print(len(filtered_sentence))
f1 = filtered_sentence
Frequecy - 1
frequency = {}
for word in filtered_sentence:
count = frequency.get(word,0)
frequency[word] = count + 1
frequency_list = frequency.keys()
for words in frequency_list:
print (words, frequency[words])
with open('machine learning.txt', 'r') as file:
data = file.read().replace('\n', '')
word_tokens = word_tokenize(data)
##print(word_tokens)
filtered_sentence = [w for w in word_tokens if not w in stop_words]
print(len(filtered_sentence))
f2 = filtered_sentence
Frequecy - 2
frequency = {}
for word in filtered_sentence:
count = frequency.get(word,0)
frequency[word] = count + 1
frequency_list = frequency.keys()
for words in frequency_list:
print (words, frequency[words])
Usage
Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.
Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.
New to Spyder? Read our tutorial
Help Variable explorer Plots Files Find
Console 1/A
service 1
robots 1
education 1
entertainment 1
applied 1
fields 1
coming 1
Together 1
changes 1
experience 2
engine 1
growth.Current 1
Uses 1
Although 1
evokes 1
thoughts 1
science 2
fiction 1
today 2
example 1
Email 2
IPython console History
conda: base (Python 3.7.6) Line 74, Col 1 UTF-8 LF RW Mem 85%


Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\Sreemanth\OneDrive\Desktop\Machine Learning\Web Mining\Lab 2\p1.py
p1.py
stop_words = (stopwords.words('english'))
stop_words.append('a')
stop_words.append('they')
stop_words.append('the')
stop_words.append('his')
stop_words.append('.')
stop_words.append(',')
stop_words.append('so')
stop_words.append('and')
stop_words.append('were')
stop_words.append('from')
stop_words.append('that')
stop_words.append('of')
stop_words.append('in')
stop_words.append('only')
stop_words.append('with')
stop_words.append('to')
word_tokens = word_tokenize(data)
filtered_sentence = [w for w in word_tokens if not w in stop_words]
print(len(filtered_sentence))
f1 = filtered_sentence
with open('machine learning.txt', 'r') as file:
data = file.read().replace('\n', '')
word_tokens = word_tokenize(data)
##print(word_tokens)
filtered_sentence = [w for w in word_tokens if not w in stop_words]
print(len(filtered_sentence))
f2 = filtered_sentence
Usage
Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.
Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.
New to Spyder? Read our tutorial
Help Variable explorer Plots Files Find
Console 1/A
designed 1
perform 2
tasks 2
within 1
domain 2
translation 1
General 1
hypothetical 1
anywhere 1
outside 1
scope 1
focuses 1
models 1
referred 1
In [32]: runfile('C:/Users/Sreemanth/OneDrive/Desktop/Machine Learning/Web Mining/Lab 2/p1.py',
wdir='C:/Users/Sreemanth/OneDrive/Desktop/Machine Learning/Web Mining/Lab 2')
511
731
In [33]:
IPython console History
conda: base (Python 3.7.6) Line 41, Col 1 UTF-8 LF RW Mem 86%