



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('Artificiaal 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
```

```
### Frequency - 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
```

```
### Frequency - 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])
```


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

```
34
35
36 filtered_sentence = [w for w in word_tokens if not w in stop_words]
37
38 print(len(filtered_sentence))
39
40 f1 = filtered_sentence
41 ## Frequency - 1
42 frequency = {}
43 for word in filtered_sentence:
44     count = frequency.get(word,0)
45     frequency[word] = count + 1
46 frequency_list = frequency.keys()
47
48 for words in frequency_list:
49     print(words, frequency[words])
50
51
52 with open('machine_learning.txt', 'r') as file:
53     data = file.read().replace('\n', '')
54
55
56 word_tokens = word_tokenize(data)
57
58 #print(word_tokens)
59 filtered_sentence = [w for w in word_tokens if not w in stop_words]
60
61 print(len(filtered_sentence))
62
63 f2 = filtered_sentence
64
65 ## Frequency - 2
66 frequency = {}
67 for word in filtered_sentence:
68     count = frequency.get(word,0)
69     frequency[word] = count + 1
70 frequency_list = frequency.keys()
71
72 for words in frequency_list:
73     print(words, frequency[words])
74
```

Usage

Here you can get help of any object by pressing **Ctrl+H** 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](#)

Console I/O

```
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
```

Python 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

```
16 stop_words = (stopwords.words('english'))
17 stop_words.append('a')
18 stop_words.append('they')
19 stop_words.append('the')
20 stop_words.append('his')
21 stop_words.append(',')
22 stop_words.append('.')
23 stop_words.append('so')
24 stop_words.append('and')
25 stop_words.append('were')
26 stop_words.append('from')
27 stop_words.append('that')
28 stop_words.append('of')
29 stop_words.append('in')
30 stop_words.append('only')
31 stop_words.append('with')
32 stop_words.append('to')
33 word_tokens = word_tokenize(data)
34
35
36 filtered_sentence = [w for w in word_tokens if not w in stop_words]
37
38 print(len(filtered_sentence))
39
40 f1 = filtered_sentence
41
42
43 with open('machine_learning.txt', 'r') as file:
44     data = file.read().replace('\n', '')
45
46
47 word_tokens = word_tokenize(data)
48
49 #print(word_tokens)
50 filtered_sentence = [w for w in word_tokens if not w in stop_words]
51
52 print(len(filtered_sentence))
53
54 f2 = filtered_sentence
55
56
```

Usage

Here you can get help of any object by pressing **Ctrl+H** 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](#)

Console I/O

```
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',
511 wdtr='C:/Users/Sreemanth/OneDrive/Desktop/Machine Learning/Web Mining/Lab 2')
511
In [33]:
```

Python console History

conda: base (Python 3.7.6) Line 41, Col 1 UTF-8 LF RW Mem 86%