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

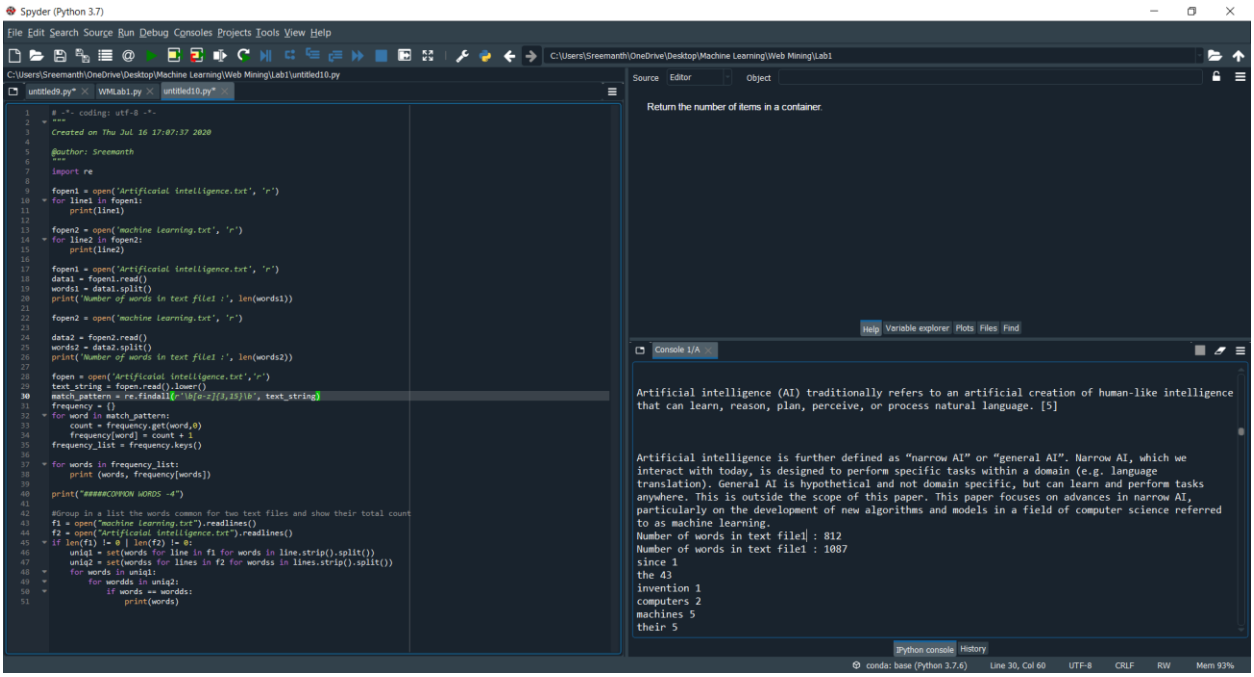
Code:

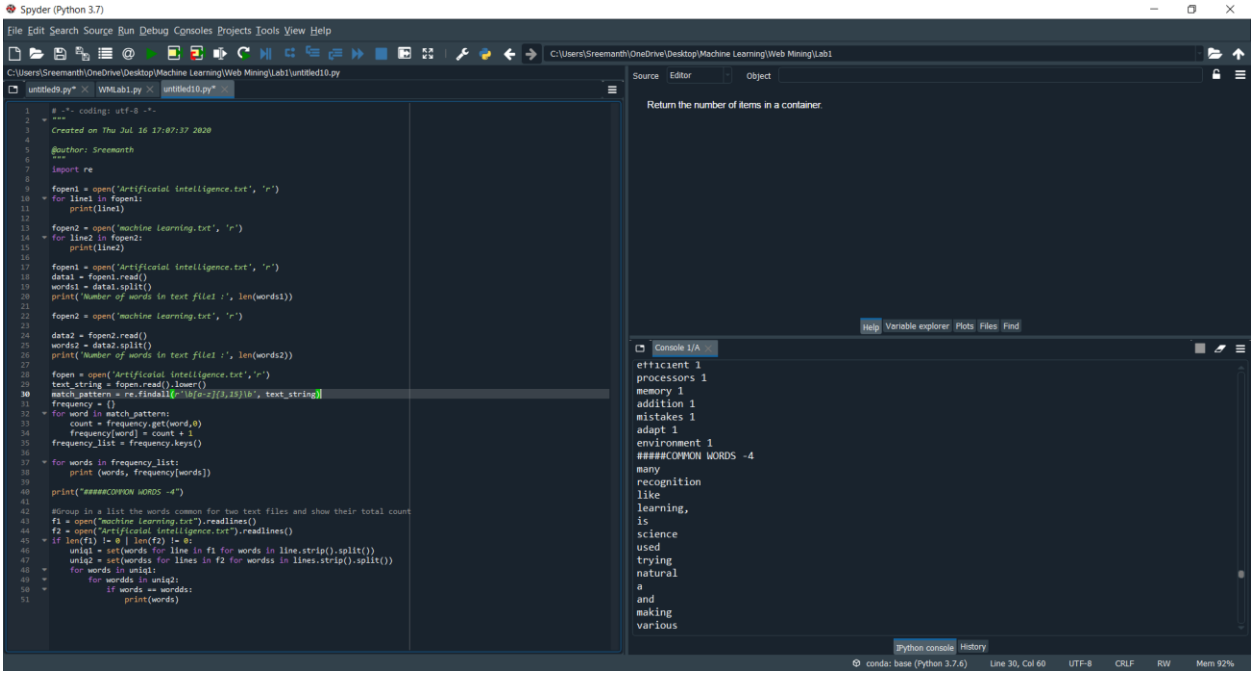
```
# -*- coding: utf-8 -*-
"""
Created on Thu Jul 16 17:07:37 2020

@author: Sreemanth
"""
import re

fopen1 = open('Artificaial intelligence.txt', 'r')
for line1 in fopen1:
    print(line1)

fopen2 = open('machine learning.txt', 'r')
for line2 in fopen2:
    print(line2)

fopen1 = open('Artificaial intelligence.txt', 'r')
```

```
data1 = fopen1.read()
words1 = data1.split()
print('Number of words in text file1 :', len(words1))

fopen2 = open('machine learning.txt', 'r')

data2 = fopen2.read()
words2 = data2.split()
print('Number of words in text file1 :', len(words2))

fopen = open('Artificaial intelligence.txt','r')
text_string = fopen.read().lower()
match_pattern = re.findall(r'\b[a-z]{3,15}\b', text_string)
frequency = {}
for word in match_pattern:
   count = frequency.get(word,0)
   frequency[word] = count + 1
frequency_list = frequency.keys()

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

print("#####COMMON WORDS -4")

#Group in a list the words common for two text files and show their total count
f1 = open("machine learning.txt").readlines()
f2 = open("Artificaial intelligence.txt").readlines()
if len(f1) != 0 | len(f2) != 0:
   uniq1 = set(words for line in f1 for words in line.strip().split())
```

```
uniq2 = set(wordss for lines in f2 for wordss in lines.strip().split())
for words in uniq1:
    for wordds in uniq2:
        if words == wordds:
            print(words)
```

Output:




Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\Sreemanth\OneDrive\Desktop\Machine Learning\Web Mining\Lab1
C:\Users\Sreemanth\OneDrive\Desktop\Machine Learning\Web Mining\Lab1\untitled10.py
untitled9.py*
WMLab1.py
untitled10.py*
-*- coding: utf-8 -*-
"""
Created on Thu Jul 16 17:07:37 2020
@author: Sreemanth
"""
import re
fopen1 = open('Artificaial intelligence.txt', 'r')
for line1 in fopen1:
print(line1)
fopen2 = open('machine learning.txt', 'r')
for line2 in fopen2:
print(line2)
fopen1 = open('Artificaial intelligence.txt', 'r')
data1 = fopen1.read()
words1 = data1.split()
print('Number of words in text file1 :', len(words1))
fopen2 = open('machine learning.txt', 'r')
data2 = fopen2.read()
words2 = data2.split()
print('Number of words in text file1 :', len(words2))
fopen = open('Artificaial intelligence.txt','r')
text_string = fopen.read().lower()
match_pattern = re.findall(r'\b[a-z]{3,15}\b', text_string)
frequency = {}
for word in match_pattern:
count = frequency.get(word,0)
frequency[word] = count + 1
frequency_list = frequency.keys()
for words in frequency_list:
print (words, frequency[words])
print("#####COMMON WORDS -4")
#Group in a list the words common for two text files and show their total count
f1 = open("machine learning.txt").readlines()
f2 = open("Artificaial intelligence.txt").readlines()
if len(f1) != 0 | len(f2) != 0:
uniq1 = set(words for line in f1 for words in line.strip().split())
uniq2 = set(words for lines in f2 for wordss in lines.strip().split())
for words in uniq1:
for wordds in uniq2:
if words == wordds:
print(words)
Source Editor
Object
Return the number of items in a container.
Help
Variable explorer
Plots
Files
Find
Console 1/A
efficient 1
processors 1
memory 1
addition 1
mistakes 1
adapt 1
environment 1
#####COMMON WORDS -4
many
recognition
like
learning,
is
science
used
trying
natural
a
and
making
various
IPython console
History
conda: base (Python 3.7.6)
Line 30, Col 60
UTF-8
CRLF
RW
Mem 92%