

**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:**





