**Code:-**

from nltk.tokenize import word\_tokenize

import nltk

from nltk.corpus import stopwords

nltk.download('stopwords')

file = open('file2.txt', encoding='utf8')

read = file.read()

file.seek(0)

print(read)

# to obtain the

# number of lines

# in file

line = 1

for word in read:

if word == '\n':

line += 1

#print("Number of lines in file is: ", line)

# create a list to

# store each line as

# an element of list

array = []

for i in range(line):

array.append(file.readline())

#print(array)

for i in range(1):

# this will convert

# the word into tokens

text\_tokens = word\_tokenize(read)

tokens\_without\_sw = [

word for word in text\_tokens ]

#print(tokens\_without\_sw)

dict = {}

uniquetokens = []

for item in tokens\_without\_sw:

if item not in uniquetokens:

uniquetokens.append(item)

for i in range(line):

print(i)

check = array[i].lower()

# print(check)

for item in uniquetokens:

if item in check:

if item not in dict:

dict[item] = []

if item in dict:

dict[item].append(i+1)

#print(dict)

print("\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*")

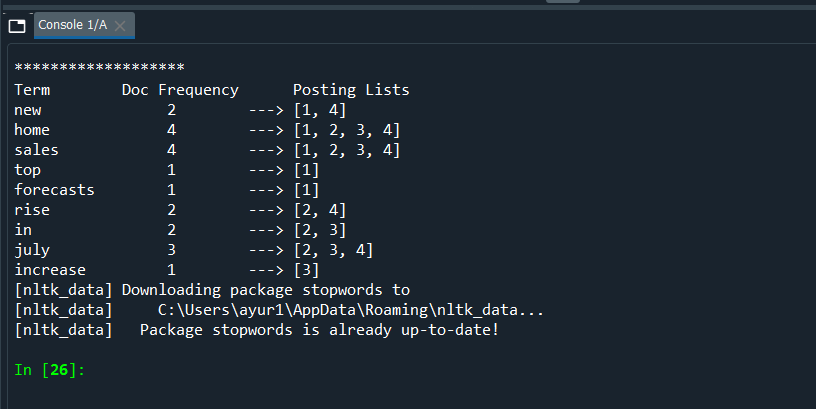
print("Term\t Doc Frequency \t Posting Lists")

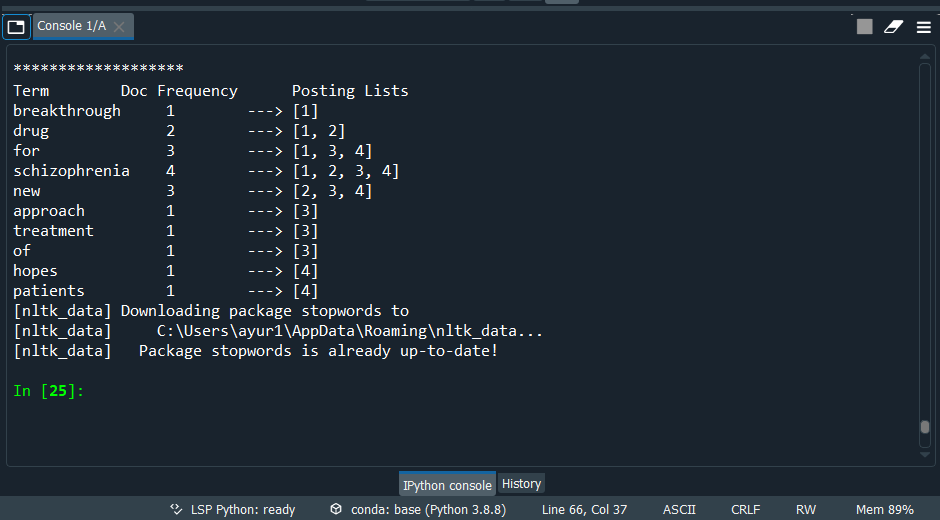
for item in dict:

# print(item," ",len(dict[item])," ",dict[item])

# print each data item.

print ("{:<16} {:<8} ---> {:<10}".format(item,len(dict[item]),str(dict[item])))





Result:-

Hence inverted index is executed for the document collection a and b.