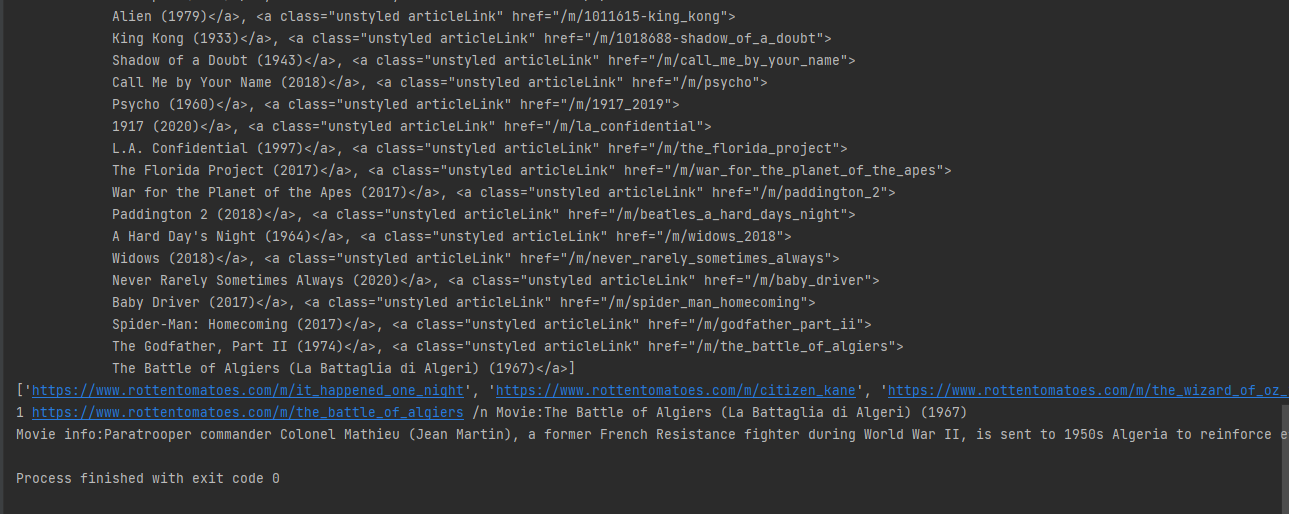
Program no:

Aim: implement a simple web crawler

Source code:

import requests  
import lxml  
from bs4 import BeautifulSoup  
url = "https://www.rottentomatoes.com/top/bestofrt/"  
headers = {  
 'User-Agent':'Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/63.0.3239.132 Safari/537.36 OPR/50.0.2762.58 (Edition Yx 01)'  
}  
f = requests.get(url, headers = headers)  
movies\_lst = []  
soup = BeautifulSoup(f.content, 'html.parser')  
movies = soup.find('table', {  
 'class':'table'  
}) .find\_all('a')  
print(movies)  
num = 0  
for anchor in movies:  
 urls = 'https://www.rottentomatoes.com' +anchor['href']  
 movies\_lst.append(urls)  
print(movies\_lst)  
num += 1  
movie\_url = urls  
movie\_f = requests.get(movie\_url, headers=headers)  
movie\_soup = BeautifulSoup(movie\_f.content, 'lxml')  
movie\_content = movie\_soup.find('div', {  
 'class':'movie\_synopsis clamp clamp-6 js-clamp'  
})  
print(num, urls, '/n', 'Movie:'+anchor.string.strip())  
print('Movie info:' + movie\_content.string.strip())

OUTPUT: 

Program:

Aim

Source code:

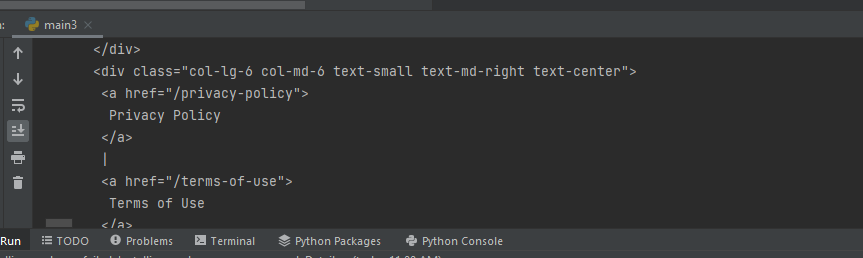
Program:

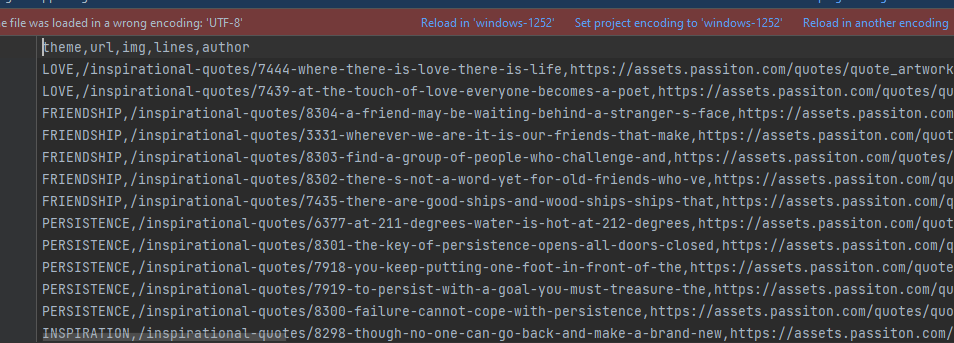
Aim

Source code:

import requests  
from bs4 import BeautifulSoup  
import csv  
  
url = "http://www.values.com/inspirational-quotes"  
r = requests.get(url)  
print("Content:")  
print(r.content)  
  
print("Prettify:")  
soup = BeautifulSoup(r.content, 'lxml')  
print(soup.prettify())  
  
quotes = []  
  
table = soup.find('div', attrs={'id':'all\_quotes'})  
  
for row in table.find\_all('div', attrs={'class':'col-6 col-lg-3 text-center margin-30px-bottom sm-margin-30px-top'}):  
 quote = {}  
 quote['theme'] = row.h5.text  
 quote['url'] = row.a['href']  
 quote['img'] = row.img['src']  
 quote['lines'] = row.img['alt'].split(" #")[0]  
 quote['author'] = row.img['alt'].split(" #")[1]  
 quotes.append(quote)  
  
filename = 'inspirational\_quotion.csv'  
with open(filename, 'w',newline='') as f:  
 w = csv.DictWriter(f, ['theme','url','img','lines','author'])  
 w.writeheader()  
 for quote in quotes:  
 w.writerow(quote)

OUTPUT:





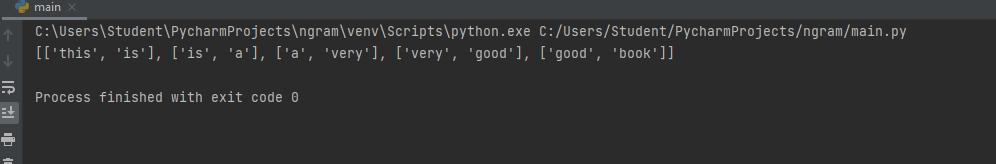
Program:

Aim:

Source Code:

def generate\_ngram(text, WordstoCombine):  
  
 words=text.split()  
 output=[]  
 for i in range(len(words) - WordstoCombine+1):  
 output.append(words[i:i + WordstoCombine])  
 return output  
x=generate\_ngram(text='this is a very good book', WordstoCombine=2)  
print(x)

Output:



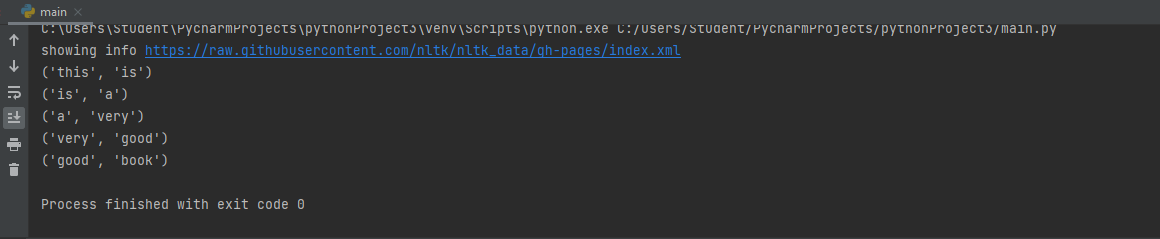
Program:

Aim: Ngram

Source Code:

import nltk  
from nltk.util import ngrams  
  
samplText ='this is a very good book'  
NGRAMS = ngrams(sequence=nltk.word\_tokenize(samplText), n=2)  
for grams in NGRAMS:  
 print(grams)

Output:



program:

Aim: python program of natural language processing for speech taging

Source Code:

import nltk

from nltk.corpus import stopwords

from nltk.tokenize import word\_tokenize, sent\_tokenize

stop\_word= set (stopwords.words('english'))

txt="Sukanya, Rajib and Naba are good friends."\

"Sukanya is getting married next year."\

"Mariage is a big step in one's life."\

"It is both exciting and frightening."

tokenized= sent\_tokenize(txt)

for i in tokenized:

wordsList = nltk.word\_tokenize(i)

wordsList = [w for w in wordsList if not w in stop\_word]

tagged = nltk.pos\_tag(wordsList)

print(tagged)

OUTPUT:

C:\Users\Student\PycharmProjects\pythonProject3\venv\Scripts\python.exe C:/Users/Student/PycharmProjects/pythonProject3/main2.py

[('Sukanya', 'NNP'), (',', ','), ('Rajib', 'NNP'), ('Naba', 'NNP'), ('good', 'JJ'), ('friends.Sukanya', 'NN'), ('getting', 'VBG'), ('married', 'VBD'), ('next', 'JJ'), ('year.Mariage', 'NN'), ('big', 'JJ'), ('step', 'NN'), ('one', 'CD'), ("'s", 'POS'), ('life.It', 'NN'), ('exciting', 'VBG'), ('frightening', 'NN'), ('.', '.')]

Process finished with exit code 0