CYCLE - 5

1. Write a program to implement a simple web crawler using Python. Extract and display the content of the page(p tag)

CODE:

```python
import requests
from bs4 import BeautifulSoup
print("SJC22MCA-2007 : ANJALA MICHAEL")
print("Batch : MCA 2022-24")
def getdata(url):
    r = requests.get(url)
    return r.content
htmldata = getdata("https://www.toppr.com/guides/essays/globalization-essay/")
soup = BeautifulSoup(htmldata,'html.parser')
data = ''
pr = len(soup.find_all('p'))
print("<P> tag ",pr)
for data in soup.find_all('p'):
    print(data.get_text())
```

OUTPUT:

```
"/home/sjcet/anjala007/python/Cycle - 5/bin/python" /home/sjcet/anjala007/sem 3/DS & ML/Cycle - 5/1.py
SJC22MCA-2007 : ANJALA MICHAEL
Batch : MCA 2022-24
<P> tag  16
Globalization refers to integration between people, companies, and governments. Most noteworthy, this integration occurs on a global scale. Furthermore, it is the process of exp

First of all, people have been trading goods since civilization began. In the 1st century BC, there was the transportation of goods from China to Europe. The goods transportatio
Globalization kept on growing gradually since 1st BC. Another significant development took place in the 7th century AD. This was the time when the religion of Islam spread. Most
True Global trade began in the Age of Discovery in the 15th century. The Eastern and Western continents were connected by European merchants. There was the discovery of America
From the 19th century, there was a domination of Great Britain all over the world. There was a rapid spread of international trade. The British developed powerful ships and trai
Finally, in 20th and 21st -Century Globalization took its ultimate form. Above all, the development of technology and the internet took place. This was a massive aid for Globali
Get the huge list of more than 500 Essay Topics and Ideas
First of all, Foreign Direct Investment (FDI) increases at a great rate. This certainly is a huge contribution of Globalization. Due to FDI, there is industrial development. Fur
Technological Innovation is another notable contribution of Globalization. Most noteworthy, there is a huge emphasis on technology development in Globalization. Furthermore, the
The quality of products improves due to Globalization. This is because manufacturers try to make products of high-quality. This is due to the pressure of intense competition. If
To sum it up, Globalization is a very visible phenomenon currently. Most noteworthy, it is continuously increasing. Above all, it is a great blessing to trade. This is because i

Which class are you in?
Download the App
Which class are you in?

Process finished with exit code 0
```

2. Write a program to implement a simple web crawler using Python. Display all hyperlinks in the page

CODE:

```python
import requests

from bs4 import BeautifulSoup

def getdata(url):

    r = requests.get(url)

    return r.content

htmldata = getdata("https://sjcetpalai.ac.in")

soup = BeautifulSoup(htmldata,'html.parser')

print("SJC22MCA-2007 : ANJALA MICHAEL")

print("Batch : MCA 2022-24")

links = soup.find_all("a")

print("Total number of links : ",len(links))

for link in links:

    if link.get("href") != "":

        print("Link :",link.get("href"),"Text :",link.string)
```

OUTPUT:

```
SJC22MCA-2007 : ANJALA MICHAEL
Batch : MCA 2022-24
Total number of links :  187
Link : https://sjcetpalai.ac.in/admissionportal/ Text : Admission 2024 - Apply Now
Link : https://sjcet.koha.sjcetpalai.ac.in/ Text : None
Link : https://sjcetpalai.ac.in/library-and-information-division/ Text : None
Link : https://www.facebook.com/SJCETPALA/ Text : Facebook
Link : https://www.instagram.com/sjcetpalai/ Text : Instagram
Link : https://www.linkedin.com/company/13462646/ Text : Linkedin
Link : https://www.youtube.com/user/SJCETPALAI Text : YouTube
Link : https://twitter.com/sjcet_palai Text : Twitter
Link : https://sjcetpalai.ac.in/ Text : None
Link : # Text : None
Link : https://sjcetpalai.ac.in Text : Home
Link : # Text : None
Link : https://sjcetpalai.ac.in/sjcet-overview/ Text : Over View
Link : https://sjcetpalai.ac.in/leadership/ Text : Leadership
Link : https://sjcetpalai.ac.in/governing-body/ Text : Governing Body
Link : https://sjcetpalai.ac.in/wp-content/uploads/2023/10/SJCET_PALAI_02-compressed.pdf Text : Organogram
Link : https://sjcetpalai.ac.in/telephone-directory/ Text : Telephone Directory
Link : https://sjcetpalai.ac.in/sjcet-palai-location/ Text : Location & Layout
Link : # Text : None
Link : https://sjcetpalai.ac.in/iqac/ Text : IQAC
Link : https://sjcetpalai.ac.in/nba-2/ Text : NBA
Link : https://sjcetpalai.ac.in/naac/ Text : NAAC
Link : https://sjcetpalai.ac.in/iso/ Text : ISO
Link : https://sjcetpalai.ac.in/sjcet-committee/ Text : Other Committees
Link : https://sjcetpalai.ac.in/policy-documents/ Text : Policy Documents
Link : # Text : None
Link : https://sjcetpalai.ac.in/admission-overview/ Text : None
```

Link : https://sjcetpalai.ac.in/admission-overview/ Text : None
Link : https://sjcetpalai.ac.in/sjcet-palai-scholarships/ Text : None
Link : https://sjcetpalai.ac.in/b-tech-admission/ Text : None
Link : https://sjcetpalai.ac.in/b-tech-admission-lateral-entry/ Text : None
Link : https://sjcetpalai.ac.in/m-tech-admission/ Text : None
Link : https://sjcetpalai.ac.in/mba-admission/ Text : None
Link : https://sjcetpalai.ac.in/mca-admission/ Text : None
Link : https://sjcetpalai.ac.in/phdprogramme/ Text : None
Link : https://sjcetpalai.ac.in/brochures/ Text : None
Link : # Text : None
Link : https://sjcetpalai.ac.in/curriculum/ Text : None
Link : https://sjcetpalai.ac.in/workshops-seminars/ Text : None
Link : https://sjcetpalai.ac.in/mentoring-and-advising/ Text : None
Link : https://sjcetpalai.ac.in/adhome/ Text : None
Link : https://sjcetpalai.ac.in/cehome/ Text : None
Link : https://sjcetpalai.ac.in/csehome/ Text : None
Link : https://sjcetpalai.ac.in/cyhome/ Text : None
Link : https://sjcetpalai.ac.in/ecehome/ Text : None
Link : https://sjcetpalai.ac.in/eshome/ Text : None
Link : https://sjcetpalai.ac.in/eeehome/ Text : None
Link : https://sjcetpalai.ac.in/mehome/ Text : None
Link : https://sjcetpalai.ac.in/mcahome/ Text : None
Link : https://sjcetpalai.ac.in/mbahome/ Text : None
Link : https://sjcetpalai.ac.in/shhome/ Text : None
Link : https://sjcetpalai.ac.in/research/ Text : None
Link : https://sjcetpalai.ac.in/office/ Text : None
Link : https://sjcetpalai.ac.in/library-and-information-division/ Text : None
Link : https://sjcetpalai.ac.in/pehome/ Text : None
Link : https://sjcetpalai.ac.in/sjcet-palai-placement/ Text : None
Link : https://sjcetpalai.ac.in/wp-content/uploads/2022/04/KTU-AC1-582-2022-8.pdf Text : None
Link : https://sjcetpalai.ac.in/sdc/ Text : None
Link : # Text : UG
Link : https://sjcetpalai.ac.in/adhome/ Text : B.Tech in Artificial Intelligence & Data Science
Link : https://sjcetpalai.ac.in/cehome/ Text : B.Tech in Civil Engineering
Link : https://sjcetpalai.ac.in/csehome/ Text : B.Tech in Computer Science & Engineering
Link : https://sjcetpalai.ac.in/cyhome/ Text : B.Tech in Computer Science & Engineering (Cyber Security)
Link : https://sjcetpalai.ac.in/ecehome/ Text : B.Tech in Electronics and Communication Engineering
Link : https://sjcetpalai.ac.in/eshome/ Text : B.Tech in Electronics & Computer Engineering
Link : https://sjcetpalai.ac.in/eeehome/ Text : B.Tech in Electrical & Electronics Engineering
Link : https://sjcetpalai.ac.in/mehome/ Text : B.Tech in Mechanical Engineering
Link : # Text : PG
Link : # Text : M.Tech
Link : https://sjcetpalai.ac.in/mehome/ Text : Advanced Manufacturing & Production Management
Link : https://sjcetpalai.ac.in/csehome/ Text : Computer Science & Engineering
Link : https://sjcetpalai.ac.in/cehome/ Text : Structural Engineering & Construction Management
Link : https://sjcetpalai.ac.in/ecehome/ Text : VLSI and Embedded Systems
Link : https://sjcetpalai.ac.in/mbahome/ Text : Masters in Business Administration (MBA)
Link : https://sjcetpalai.ac.in/mcahome/ Text : Masters in Computer Applications (MCA)
Link : # Text : PhD
Link : https://sjcetpalai.ac.in/csehome/ Text : Computer Science & Engineering
Link : https://sjcetpalai.ac.in/ecehome/ Text : Electronics & Communication Engineering
Link : https://sjcetpalai.ac.in/mehome/ Text : Mechanical Engineering
Link : https://sjcetpalai.ac.in/mbahome/ Text : Management
Link : https://sjcetpalai.ac.in/mcahome/ Text : Computer Applications
Link : # Text : None
Link : https://sjcetpalai.ac.in/raerest/ Text : International Conference
Link : https://sjcetpalai.ac.in/sjcet-redhat-academy/ Text : Red Hat® Academy
Link : https://sjcetpalai.ac.in/sports/ Text : Sports
Link : https://sjcetpalai.ac.in/sargam/ Text : Arts & Culture (Sargam)
Link : https://sjcetpalai.ac.in/tedxsjcetpalai/ Text : TEDxSJCETPalai
Link : https://sjcetpalai.ac.in/campus-events/ Text : Campus Events
Link : https://sjcetpalai.ac.in/professional-societies/ Text : Professional Societies

Link : https://sjcetpalai.ac.in/sjcet-industrial-consultancy Text : Industrial Consultancy
Link : https://sjcetpalai.ac.in/ncc-sjcet/ Text : National Cadet Corps
Link : https://sjcetpalai.ac.in/nss-sjcet/ Text : National Service Scheme
Link : https://sjcetpalai.ac.in/women-development/ Text : Women in Development cell
Link : https://sjcetpalai.ac.in/karate-club/ Text : Karate Club
Link : # Text : None
Link : https://sjcetpalai.ac.in/library-and-information-division/ Text : College Library
Link : https://sjcetpalai.ac.in/residential-facilities/ Text : Residential Facilities
Link : https://sjcetpalai.ac.in/bank-atm-and-postal-services/ Text : Banking & Postal Services
Link : https://sjcetpalai.ac.in/transportation-facility/ Text : Transportation Facilities
Link : https://sjcetpalai.ac.in/medicare-sjcet/ Text : Medicare
Link : https://sjcetpalai.ac.in/student-counseling/ Text : Student Counseling
Link : # Text : None
Link : https://swayam.gov.in/ Text : Swayam Portal
Link : http://pay.sjcetpalai.ac.in/ Text : Online Fee Payment
Link : https://sjcetpalai.etlab.in/ Text : etlab SJCET
Link : https://sjcetpalai.hrm.etlab.in/ Text : etlab HRM
Link : # Text : SJCET LMS
Link : https://celms.sjcetpalai.ac.in/ Text : Dept. of Civil Engineering
Link : https://cse.lms.sjcetpalai.ac.in/ Text : Dept. of Computer Science & Engineering
Link : https://ecelms.sjcetpalai.ac.in/ Text : Dept. of Electronics & Communication Engineering
Link : https://eeelms.sjcetpalai.ac.in/ Text : Dept. of Electrical & Electronics Engineering
Link : https://aeilms.sjcetpalai.ac.in/ Text : Dept. of Electronics & Instrumentation Engineering
Link : https://melms.sjcetpalai.ac.in/ Text : Dept. of Mechanical Engineering
Link : http://117.254.186.104/mca_lms/ Text : Master of Computer Applications
Link : https://mbalms.sjcetpalai.ac.in/ Text : Mater of Business Administration
Link : https://shlms.sjcetpalai.ac.in/ Text : Science & Humanities Department
Link : https://ailms.sjcetpalai.ac.in/ Text : Dept. of Artificial Intelligence & Data Science
Link : https://ecslms.sjcetpalai.ac.in/ Text : Dept. of Electronics & Computer Engineering
Link : https://cscs.lms.sjcetpalai.ac.in/ Text : Dept. of Computer Science & Engineering (Cyber Security)

Link : https://sjcetpalai.ac.in/wp-content/uploads/2022/11/EOA_Report_22-23.pdf Text : EOA Report 2022-2023
Link : https://sjcetpalai.ac.in/wp-content/uploads/2023/06/EOA-Report-2023-24.pdf Text : EOA Report 2023-2024
Link : https://sjcetpalai.ac.in/wp-content/uploads/2023/10/Mandatory-Disclosure-Updated-as-on-August-2023-Final-.pdf Text : Mandatory Disclosure
Link : https://sjcetpalai.ac.in/financialreports/ Text : Auditors Reports
Link : https://sjcetpalai.ac.in/wp-content/uploads/2022/04/SJCET-STAFF-POLICY-compressed.pdf Text : Staff Policy
Link : https://sjcetpalai.ac.in/wp-content/uploads/2023/01/St.-Josephs-College-of-Engg.-_-Tech.-Kottayam-Kerala26_8_2022_17_11_8.pdf Text : NBA Accreditation Certificate ECE &
Link : https://sjcetpalai.ac.in/wp-content/uploads/2022/06/NBA-2022-CS-and-CE.pdf Text : NBA Accreditation Certificate CSE & CE
Link : https://sjcetpalai.ac.in/wp-content/uploads/2020/12/HoD-2020.docx Text : HOD
Link : https://sjcetpalai.ac.in/wp-content/uploads/2021/03/Appraisal-360-1.docx Text : Faculty
Link : https://sjcetpalai.ac.in/wp-content/uploads/2020/12/Non-Teaching-Staff-2020.docx Text : Staff
Link : https://sjcetpalai.ac.in/wp-content/uploads/2021/03/AICTE-Pay-Revision-Kerala.16.2.211.pdf Text : Calculation Of 360 Feedback Score
Link : https://sjcetpalai.etlab.in/user/login Text : etlab SJCET
Link : https://sjcetpalai.hrm.etlab.in/backend/web/site/login Text : etlab HRM
Link : https://lms.sjcetpalai.ac.in Text : Learning Management System
Link : https://mail.google.com/a/sjcetpalai.ac.in Text : Campus Mail
Link : http://gms.sjcetpalai.ac.in/sjcet Text : Grievance Redressal
Link : http://elective.sjcetpalai.ac.in/ Text : Elective Portal
Link : https://sjcetpalai.ac.in/careers-sjcet-palai/ Text : Careers@SJCET
Link : https://www.instagram.com/sjcet_palai/ Text : Image Gallery
Link : https://www.youtube.com/user/SJCETPALAI/videos Text : Video Gallery
Link : https://sjcetpalai.ac.in/publications-sjcet-palai/ Text : Publications
Link : https://forms.gle/L4B8UFeJNk4R1rte9 Text : Feedback on Curriculum/Institute from Teachers
Link : https://sjcetpalai.ac.in/sjcet-palai-alumni-association-sjcetpaan/ Text : SJCETPAAN
Link : https://www.facebook.com/SJCETPALA/ Text : Facebook
Link : https://www.instagram.com/sjcetpalai/ Text : Instagram
Link : https://www.linkedin.com/company/13462646/ Text : Linkedin
Link : https://www.youtube.com/user/SJCETPALAI Text : YouTube
Link : https://twitter.com/sjcet_palai Text : Twitter
Link : None Text : None
Link : None Text : X

3. Program for Natural Language Processing which performs n-grams(without using library)

CODE:

```python
print("SJC22MCA-2007 : ANJALA MICHAEL")

print("Batch : MCA 2022-24")

def gen_ngrams(text, wordsToCombine):

    words = text.split()

    output = []

    for i in range(len(words)-wordsToCombine+1):

        output.append(words[i:i+wordsToCombine])

    return output


x = gen_ngrams(text='Using the iris data set, implement the KNN algorithm',wordsToCombine=3)

print(x)
```

OUTPUT:

```
SJC22MCA-2007 : ANJALA MICHAEL
Batch : MCA 2022-24
[['Using', 'the', 'iris'], ['the', 'iris', 'data'], ['iris', 'data', 'set,'],




['data', 'set,', 'implement'], ['set,', 'implement', 'the'], ['implement', 'the', 'KNN'],




['implement', 'the', 'KNN'], ['the', 'KNN', 'algorithm']]
```

4. Program for Natural Language Processing which performs n-grams(using nltk library)

CODE:

```python
print("SJC22MCA-2007 : ANJALA MICHAEL")
print("Batch : MCA 2022-24")
from nltk import ngrams
sentence = 'I reside in India'
n = 3
trigrams = ngrams(sentence.split(),n)
for grams in trigrams:
    print(grams)
```

OUTPUT:

```
"/home/sjcet/anjala007/python/Cycle - 5/bin/python" /home/sjcet/anj
SJC22MCA-2007 : ANJALA MICHAEL
Batch : MCA 2022-24
('I', 'reside', 'in')
('reside', 'in', 'India')

Process finished with exit code 0
```

5. For given text,

  - perform word
  - sentence tokenization
  - Remove the stop words from the given text
  - create n-grams

CODE:

```python
import nltk

from nltk import ngrams

from nltk.corpus import stopwords

nltk.download('punkt')

from nltk.tokenize import sent_tokenize,word_tokenize

text1 = "The data given satisfies the requirement for model generation. This
is used in Data Science Lab"

print("Sentance tokenization : ")

print(sent_tokenize(text1))

print("Word tokenization : ")

print(word_tokenize(text1))

text = word_tokenize(text1)

text2 = [word for word in text if word not in stopwords.words('english')]

print("")

print("Removing stop words : ")

print(text2)

print("")

print("n grams : ")

unigrams = ngrams(text2,3)

for grams in unigrams:

    print(grams)
```

OUTPUT:

```
"/home/sjcet/anjala007/python/Cycle - 5/bin/python" /home/sjcet/anjala007/sem 3/DS & ML/Cycle - 5/5.py
[nltk_data] Downloading package punkt to /home/sjcet/nltk_data...
Sentance tokenization :
['The data given satisfies the requirement for model generation.', 'This is used in Data Science Lab']
Word tokenization :
['The', 'data', 'given', 'satisfies', 'the', 'requirement', 'for', 'model', 'generation', '.', 'This', 'is', 'used', 'in', 'Data', 'Science', 'Lab']

Removing stop words :
['The', 'data', 'given', 'satisfies', 'requirement', 'model', 'generation', '.', 'This', 'used', 'Data', 'Science', 'Lab']

n grams :
('The', 'data', 'given')
('data', 'given', 'satisfies')
('given', 'satisfies', 'requirement')
('satisfies', 'requirement', 'model')
('requirement', 'model', 'generation')
('model', 'generation', '.')
('generation', '.', 'This')
('.', 'This', 'used')
('This', 'used', 'Data')
('used', 'Data', 'Science')
('Data', 'Science', 'Lab')
[nltk_data]   Package punkt is already up-to-date!
```

6. Given dataset contains 200 records and five columns, two of which describe the customer's annual income and spending score. The latter is a value from 0 to 100. The higher the number, the more this customer has spent with the company in the past:

Using k means clustering creates 6 clusters of customers based on their spending pattern.

- Visualize the same in a scatter plot with each cluster in a different color scheme.
- Display the cluster labels of each point.(print cluster indexes)
- Display the cluster centers.
- Use different values of K and visualize the same using scatter plot

CODE:

```python
import pandas as pd


customer = pd.read_csv('customer_data.csv')

customer.head()

import matplotlib.pyplot as plt


point = customer.iloc[:,3:5].values

x = point[:,0]

y = point[:,1]

plt.scatter(x,y,s=50,alpha=0.7)

plt.xlabel('Annual income (k$)')

plt.ylabel('Spending Score')

plt.show()



from sklearn.cluster import KMeans


kmeans = KMeans(n_clusters=6,random_state=0)

kmeans.fit(point)

predicted_cluster_indexes = kmeans.predict(point)
```

```python
plt.scatter(x,y,c=predicted_cluster_indexes,s=50,alpha=0.7,cmap='viridis')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score')
plt.show()


from sklearn.cluster import KMeans


kmeans = KMeans(n_clusters=7,random_state=0)
kmeans.fit(point)
predicted_cluster_indexes = kmeans.predict(point)
plt.scatter(x,y,c=predicted_cluster_indexes,s=50,alpha=0.7,cmap='viridis')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score')
plt.title('Cluster centers')
plt.show()
```

OUTPUT:

Cluster centers