# Web mining Assignment -3

Note: You need to use "Beautiful Soup" package and "Requests" module. Programs need to be written in Python language 3.8 version. Use of Regular expressions allowed [re module]

**NAME: GEROSH GEORGE** 

**REG NO: 19BCE1403** 

- 1. Web page: <a href="https://www.vit.ac.in">https://www.vit.ac.in</a>
  - Print the "title" of the page
  - Print out all the anchor tags with the class = "nav-link"

```
import requests
from bs4 import BeautifulSoup

url=input('Enter the url: ')

resp = requests.get(url)

page = resp.text
soup= BeautifulSoup(page, 'html.parser')

print(f'Title of page: {soup.title.text}')

a_tags = soup.findAll(name='a', attrs={'class':'nav-link'})

for a in a_tags:
    print(a)
```

```
(a class="nav-link" href="https://vit.ac.in/programmes-offered-1" title="Programmes Offered">Programmes Offered(/a)
(a class="nav-link" href="https://vit.ac.in/programmes-offered-2021-22" title="AY 2021-22">AY 2021-22(a)
(a class="nav-link" href="https://vit.ac.in/programmes-offered-2020-21" title="AY 2020-21">AY 2020-21*(a)
(a class="nav-link" href="https://vit.ac.in/schools" title="Curriculum">Curriculum*(/a)
(a class="nav-link" href="https://vit.ac.in/academics/fics" title="Fics">FFCS</a>(a class="nav-link" href="https://vit.ac.in/academics/library" title="Library">Library*(a)
(a class="nav-link" href="https://vit.ac.in/academics/library" title="Library">Library*(a)
(a class="nav-link" href="https://vit.ac.in/academics/fics" title="Fics">Library*(a)
(a class="nav-link" href="https://vit.ac.in/academics-feedback" title="loops">Library*(a)
(a class="nav-link" href="https://vit.ac.in/academics-feedback">Library*(a)
(a class="nav-link" href="https://vit.ac.in/admissions/overview" title="Overview">Doverview*(a)
(a class="nav-link" href="https://vit.ac.in/admissions/overview" title="Overview">Doverview*(a)
(a class="nav-link" href="https://vit.ac.in/admissions/overview" title="Programmes Offered">Programmes Offered*(a)
(a class="nav-link" href="https://vit.ac.in/admissions/reparammes-offered" title="Programmes Offered">Programmes Offered*(a)
(a class="nav-link" href="https://vit.ac.in/admissions/research/Integrated Ph.D" title="Programmes Offered">Programmes Offered*(a)
(a class="nav-link" href="https://vit.ac.in/admissions/research/Integrated Ph.D" title="Integrated Ph.D">Integrated Ph.D">In
```

```
<a class="nav-link" href="https://vit.ac.in/academics-more/Contact us" title="Contact us">Contact us</a>
       <a class="nav-link" href="https://vit.ac.in/research" title="Research">Research()a>
<a class="nav-link" href="https://vit.ac.in/research" title="Overview">Overview</a>
<a class="nav-link" href="https://vit.ac.in/research/academic" title="Academic Research">Academic Research</a>
     <a class="nav-link" href="https://vit.ac.in/research/academic" title="Academic Research">Academic Research(a)
<a class="nav-link" href="https://vit.ac.in/research/sonsored-research">Sponsored Research">Sponsored Research</a>
<a class="nav-link" href="https://vit.ac.in/iprcell" title="IPR Cell">IPR Cell</a>
<a class="nav-link" href="https://vit.ac.in/research/centers-list" title="Research Centers">Research Centers</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/overview" title="Campus Life">Campus Life</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/overview" title="Overview">Overview</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/fests" title="Fests">Fests</a>
<a class="nav-link" href="https://vit.ac.i

<a class="nav-link" href="https://vit.ac.in/campuslife/fstudentswelfare" title="Students' Welfare">Students' Welfare(a>
<a class="nav-link" href="https://vit.ac.in/campuslife/studentswelfare" title="Student Clubs">Student Clubs</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/Clubs" title="Student Clubs">Student Clubs</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/Chapters" title="Student Chapters">Student Chapters</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/Chapters" title="Student Chapters">Student Chapters</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/Campus-Events" title="Campus Events">Campus Events</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/Counselling-Division" title="Counselling Division">Counselling Division</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/grievancecell" title="General Grievance Redressal Committee">General Grievance Redressal Committee</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/grievancecell" title="General Grievance Redressal Committee">General Grievance Redressal Committee</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/grievancecell" title="General Grievance Redressal Committee">General Grievance Redressal Committee</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/grievancecell" title="General Grievance Redressal Committee">General Grievance Redressal Committee</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/grievancecell" title="General Grievance Redressal Committee">General Grievance Redressal Committee</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/grievancecell" title="General Grievance Redressal Committee">General Grievance Redressal Committee</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/grievancecell" title="General Grievancecell" title="general Grievancecell" https://wit.ac.in/campus-category/grievancecell" title="general Grievancecell" https://wit.ac.in/campus-c
     Redressal Committee</a>
<a class="nav-link" her="https://vit.ac.in/campus-category/Student-Council" title="Student Council">Student Council</a>
<a class="nav-link" href="https://vit.ac.in/campus-category/Student-Council" title="Student Council">Student Council</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/sports" title="Library">Library</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/sports" title="Sports">Sports</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/hostels" title="Hostels">Hostels</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/hostels" title="Hostels"
<a class="nav-link" href="https://vit.ac.in/campuslife/hostels" title="Hostels" hostels</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/hostels" title="Hostels" hostels</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/hostels" href="https://vit.ac.in/campuslife/hostels" href="https://vit.ac.in/campuslife/hostels" href="https://vit.ac.in/campuslife/hostels" href="https://vit.ac.in/campuslife/hostels" href="https://vit.ac.in/campuslife/hostels" href="https://vit.ac.in/campuslife/h
          <a class="nav-link" href="https://vit.ac.in/campuslife/healthservices" title="Health Services">Health Services</a>
<a class="nav-link" href="https://vit.ac.in/campuslife/otheramenities" title="Other Amenities">Other Amenities</a>
       <a class="nav-link" nref="https://vit.ac.in/campuslire/orneramenties" title="Utner Amenities" byther Amenities" byther Amenities" byther Amenities "byther Amenities "byther Amenities" byther Amenities "byther Amenities "byther Amenities" byther Amenities "byther Amenities "byther Amenities "byther Amenities "byther Amenities" byther Amenities "byther Byther "byther Amenities "byther Byther "byther Byther "byther "byther "byther "byther "byther "byther "byther "byther "by
       <a class="nav-link" nref="nttps://vit.ac.in/transcripts-alumni">\text{Transcripts for Alumni</a>
<a class="nav-link" href="https://vit.ac.in/transcripts-alumni">\text{Transcripts for Alumni</a>
<a class="nav-link" href="https://vit.applydirect.org/student" target="_blank">\text{MES} attestation</a>
<a class="nav-link">\text{MES} attestation</a
     <a class="nav-link" href="https://vit.applydirect.org/student" target="_blank">WES attestation</a>
<a class="nav-link" href="https://applications.wes.org/required-documents/Search/Result?cId=2&amp;eId=38&amp;iNm=Vellore%20Institute%20of%2
0Technology%20%28VIT%29&amp;cNm=Bachelor%20of%20Applied%20Arts&amp;iSec=University%20Education&amp;iSub=Bachelor%27s%20Degree%20Programs&am
p;comp=1&amp;isEca=False" target="_blank">WES Application Process</a>
<a class="nav-link" href="https://www.wes.org/advisor-blog/send-electronic-academic-records/" target="_blank">WES Advisory Blog</a>
<a colspan="2">WES Advisory Blog</a>
<a colspan="2
       <a class="nav-link" href="https://vit.ac.in/files/e-SANAD-Notification.pdf" target="_blank">e-Sanad</a>
<a class="nav-link" href="https://web.vit.ac.in/CertificateVerification/login" target="_blank">Certificate Verification for Registered Agen
     cles(/a)
{a class="nav-link" href="https://vit.ac.in/centers/asc" target="_blank">Academic Staff College</a>
{a class="nav-link" href="http://www.vittbi.com/#/" target="_blank">VIT TBI</a>
{a class="nav-link" href="javascript:void(0)">Parents</a>
{a class="nav-link" href="https://vtop.vit.ac.in/" target="_blank">Parent Login</a>
{a class="nav-link" href="https://vtop.vit.ac.in/" target="_blank">Parent Login</a>

       {a class="nav-link" href="https://vit.ac.in/campus-category/Counselling-Division">Counselling Division </a>
<a class="nav-link" href="https://vit.ac.in/guest-house" target="_blank">Guest House</a>
<a class="nav-link" href="https://vit.ac.in/guest-house" target="_blank">Guest House</a>
<a class="nav-link" href="https://vit.ac.in/redressal">Grievance Redressal</a>
<a class="nav-link">Grievance Redressal</a>
<a class="nav-link">Grievance Redressal</a>
<a class
     <a class="nav-link" href="https://vit.ac.in/redressal">Grievance Redressal</a>
<a class="nav-link" href="https://vit.ac.in/hotels-in-vellore" target="_blank">Hotels in Vellore</a>
<a class="nav-link" href="javascript:void(0)">Students</a>
<a class="nav-link" href="https://mail.google.com/" target="_blank">VIT Gmail</a>
<a class="nav-link" href="https://vtop.vit.ac.in/" target="_blank">Student Login</a>
<a class="nav-link" href="https://vtop.vit.ac.in/" target="_blank">Student Login</a>
<a class="nav-link" href="https://vit.ac.in/" target="_blank">Student Login</a>
<a class="nav-link" href="https://vit.ac.in/sites/default/files/SBST_Freshers_App.rar" target="_blank">SBST_Freshers_App</a>
<a class="nav-link" href="https://vit.ac.in/anti-ragging-committee" target="_blank">Anti Ragging Committee</a>
<a class="nav-link" href="https://vit.ac.in/capability-enhancement-scheme" target="_blank">Capability_Enhancement_Schemes</a>
<a class="nav-link" href="https://vit.ac.in/sites/default/files/FormatGuidelines.doc" target="_blank">Capability_Enhancement_Scheme</a>
<a class="nav-link" href="https://vit.ac.in/sites/default/files/FormatGuidelines.doc" target="_blank">Capability_Enhancement_Scheme</a>
<a class="nav-link" href="https://vit.ac.in/sites/default/files/FormatGuidelines.doc" target="_blank">Capability_Enhancement_Scheme</a>
<a class="nav-link" href="https://vit.ac.in/sites/default/files/FormatGuidelines.doc" target="_blank">Capability_Enhancement_Scheme</a>
<a class="nav-li
       <a class="nav-link" href="https://vit.ac.in/sites/default/files/FormatGuidelines.doc"
<a class="nav-link" href="https://vit.ac.in/internal-complaints-committee">ICCC/a>
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            target="_blank">Industrial Visit</a>
<a class="nav-link" href="https://vit.ac.in/internal-complaints-committee">ICC</a>
<a class="nav-link" href="javascript:void(0)">Alumni</a>
<a class="nav-link" href="pavascript:void(0)">Alumni</a>
<a class="nav-link" href="https://vit.ac.in/alumni_progression" target="_blank">Alumni Progression</a>
<a class="nav-link" href="https://vit.ac.in/instruction">Medium of Instruction / Migration Certificate</a>
<a class="nav-link" href="https://vit.ac.in/op/initialProcess" target="_blank">Op/instruction / Migration Certificate</a>
<a class="nav-link" href="https://vit.ac.in/op/initialProcess" target="_blank">Op/instruction / Migration Certificate</a>
<a class="nav-link" href="https://vit.ac.in/alumni-evonts" target="_blank">Alumni - Evonts</a>
<a class="nav-link" href="https://vit.ac.in/detailview/alumni-photo-gallery" target="_blank">Alumni - Photo Gallery</a>
<a class="nav-link" href="https://wit.ac.in/alumni-office-contact" target="_blank">Alumni contact</a>
<a class="nav-link" href="https://vit.ac.in/alumni-office-contact" target="_blank">Alumni contact</a>
<a class="nav-link" href="https://campustour.vit.ac.in/" target="_blank">Campus Tour</a>
<a class="nav-link" href="https://wit.ac.in/alumni-office-contact" target="_blank">Alumni contact</a>
<a class="nav-link" href="https://campustour.vit.ac.in/" target="_blank">Campus Tour</a>
<a class="nav-link" href="https://wit.ac.in/alumni-office-contact" target="_blank">Campus Tour</a>
<a c
```

<a class="nav-link" href="https://vit.ac.in/internationalrelations/sap" title="SAP">SAP</a>
<a class="nav-link" href="https://vit.ac.in/admissions/international/overview" title="International Admissions">International Admissions</a</pre>

- 2. Web page: <a href="https://vit.ac.in/school/allfaculty/site/computer-applications">https://vit.ac.in/school/allfaculty/site/computer-applications</a>
  - a. Print out all the faculty names using the class id "title2" and their research area by devising appropriate algorithm you need to use methods of beautifulsoup
  - b. Find the facebook link, twitter link, instagram linkand linkedin link of of VIT from the page content and identify the class names given for each.
  - c. List out the DOM hierarchy of the page [Find all the children and the relationship between the children

```
import requests
from bs4 import BeautifulSoup
url=input('Enter the url: ')
resp = requests.get(url)
page = resp.text
soup= BeautifulSoup(page, 'html.parser')
names = []
research=[]
print("Faculty names along with their research area: ")
for h3 in soup.findAll('h3',{'class':'title2'}):
    parent = h3.parent
    p_tags = parent.find_all('p')
    if len(p_tags) == 3:
        if len(p_tags[2].text)==1:
            research.append('None')
        else:
            research.append(p_tags[2].text)
    else:
        research.append('None')
    names.append(h3.text)
for faculty,research in zip(names,research):
    print(f'{faculty}: {research}')
print("\nSocial Links: ")
links_span = soup.find('span',{'class':'soclia_links'})
links = links_span.find_all('a')
for link in links:
    print(f"{link['title']} : {link['href']}")
   print(f"class name: {link['class']}\n")
```

a)

```
(venv) PS C:\Users\Gerosh Shibu George\Desktop\VIT College\Third Year\Web Minning\Programs> python exercise2.py
Enter the url: https://vit.ac.in/school/allfaculty/site/computer-applications
Faculty names along with their research area:
Dr.Ramkumar T: Data Mining & Big Data Analytics
Dr. Ephzibah E.P: Data Mining and Artificial Intelligence
Dr. Karthikeyan P: Cloud computing, Web Services
Dr. Manivannan S.S: Network and Information Security, IoT and Machine Learning Dr. Meenatchi S: Computer Science Hardware and Architecture
Dr.Senthil Murugan B: None
Dr.Shynu P.G: Cloud computing, Information Security, Data Science
Dr. Uma Maheswari G: Computer Science Information Systems
Dr.Venketesh P: None
Dr.Deepa N: Predictive Analytics
Dr.Jayalakshmi P: Networks
Dr. Senthil Kumar N: Semantic Web; Information Retrieval
Ms. Deepa P: None
Ms. Manisha R. Patil: None
Ms. Manjupriya R: None
Mr. Ravindran U: None
Mr.Sreeraag G: None
```

b)

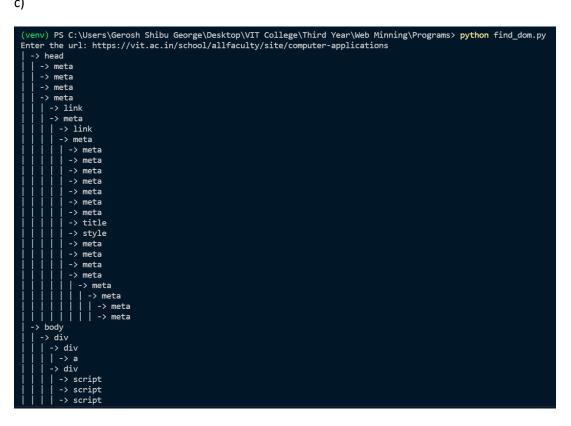
```
Social Links:
facebook : https://www.facebook.com/VITuniversity/
class name: ['face_book_icon', 'f_icon_be']

twitter : https://twitter.com/vit_univ
class name: ['twitter_icon', 'f_icon_be']

LinkedIn : https://www.linkedin.com/school/vellore-institute-of-technology/
class name: ['linkedin_icon', 'f_icon_be']

youtube : https://www.youtube.com/c/VITUniversityVellore
class name: ['youbtube_icon', 'f_icon_be']

instagram : https://www.instagram.com/vellore_vit/
class name: ['insta_icon', 'f_icon_be']
```



```
-> section
 -> div
    -> div
      -> div
        -> div
          -> div
            -> div
              -> a
               -> img
          -> div
             -> div
               -> ul
               | -> li
               -> ul
                 -> li
                 -> li
                span
               -> div
                -> div
  header
  -> div
   -> div
    | -> img
    -> div
    -> div
      -> div
        -> div
          -> div
```

| -> div<br>                  -> div<br>                    -> div<br>                  -> div<br>                    -> span<br>                    -> li  |  |
|---|--|
| -> div  |  |
| -> div<br>                    -> div<br>                      -> div  |  |
| -> div  |  |
| -> div  |  |
|   |  |
| -> ul   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
| -> div<br>                -> div<br>                -> div<br>                -> div<br>                -> span<br>                  -> ul<br>        -> div<br>          -> div<br>            -> div<br>              -> div<br>                -> li<br>                  -> a<br>                  -> a<br>                  -> a<br>                  -> li<br>                  -> a<br>                  -> li<br>                  -> li<br>                  -> li |  |
| -> div  |  |
| -> div  |  |
| -> ul   |  |
| -> 1i   |  |
|   |  |
| -> div<br>                  -> div<br>                  -> ul<br>                    -> li<br>                      -> li<br>                        -> li<br>                      -> li   |  |
|   |  |
| -> 11   |  |
|   |  |
| -> li<br>              -> div   |  |
|   |  |
|   |  |
|   |  |
| -> div  |  |
| -> div<br>                      -> div<br>                    -> div  |  |
| -> div  |  |
| -> div  |  |
| -> script   |  |
|   |  |
|   |  |
|   |  |
| -> div<br>                -> div<br>                  -> div<br>                    -> div<br>                    -> div<br>            -> script<br>            -> div<br>            -> div   |  |
|   |  |

|          |       |                          | -> div   |
|----------|-------|--------------------------|--|
| - ! !    | !!    | !!!!                     | -> div   |
| _! !     | !!    | !!!!                     |  |
| -        |       |                          |  |
|          |       | $  \cdot   \cdot  $      |  |
| -1       | -1    | $I \mid I \mid I \mid I$ |  |
| - 1 1    |       |                          |  |
| Ιİ       | Ιİ    | iiii                     |  |
| i i      |       | iiii                     | -> h3<br>        -> ul<br>      -> div   |
| i i      | i i   | iiii                     |  |
| i i      |       | iiii                     |  |
| - 1 1    | - 1 1 | 111                      |  |
| -        |       |                          |  |
| -        | -     |                          | -> div     -> div       -> div       -> h3         -> div       -> h3         -> ul       -> h3         -> ul       -> h3         -> h3         -> h3         -> h3         -> h3         -> h3         -> div |
| -        | -     | ! ! ! !                  |  |
| -        | !!    | !!!!                     | -> u1  |
| - !!     | !!    | !!!!                     | -> div   |
| !!       | !!    | !!!!                     | -> h3<br>        -> ul   |
| <u> </u> | -     |                          |  |
| -        | -     | $  \cdot   \cdot  $      |  |
|          |       |                          |  |
|          |       |                          | -> div   |
| -1       | -1    |                          | -> div   |
|          | -1    |                          | -> div<br>    -> div<br>      -> ul  |
| ΙÌ       | Ιİ    | iiii                     |  |
| i i      | i i   | iiii                     |  |
| i i      | i i   |                          |  |
| i i      | i i   |                          | -> li<br>      -> span<br>        -> a<br>        -> a<br>        -> a<br>        -> a   |
| i i      | - i i | 111                      |  |
| - 1 1    | - 1 1 | 111                      |  |
| -        | -     |                          |  |
| -        | -     | ! ! ! !                  | -> a   |
| -        | - ! ! | !!!!                     |  |
| - !!     | !!    | !!!!                     | -> hr  |
| ļļ       | i i   |                          | -> div   |
|          | Ιİ    |                          |  |
|          | ļļ.   | İİİİ                     |  |
|          |       |                          |  |

| iiii                | -> a        |
|---------------------|-------------|
| iiii                | -> i        |
| 111                 | -> div      |
|                     | -> script   |
| $  \cdot  $         | -> script   |
| $  \cdot  $         | -> script   |
| $  \cdot  $         | -> script   |
| $  \cdot   \cdot  $ | -> script   |
| $  \cdot   \cdot  $ | -> script   |
|                     | -> script   |
| $  \cdot   \cdot  $ | -> script   |
| $  \cdot   \cdot  $ | -> link     |
| $  \cdot   \cdot  $ | -> noscript |
|                     | -> link     |
|                     | -> div      |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     | -> script   |
|                     |             |

3.

Web page: https://sermitsiaq.ag/english

- a. Find all the items of class="menu" and print out the items of the menu with class names With class names "first leaf", "leaf" and "last leaf".
- b. Find all items with ids containing string "menu" in them
- c. Find all items with tag "article".
- d. List out the DOM hierarchy of the page

```
import requests
from bs4 import BeautifulSoup as bs
import re
url=input('Enter the url: ')
resp = requests.get(url,headers={'User-Agent': 'Mozilla/5.0'})
page = resp.text
soup= bs(page,'html.parser')
items = soup.findAll('ul',{'class':'menu'})
print(f"Number of items of class menu: {len(items)}")
print("\nItems inside class Menu with classes 'first leaf', 'leaf' and 'last l
eaf'")
for item in items:
    elements = item.findAll('li',{'class':['leaf','last leaf','first leaf']})
   for elem in elements:
        print(elem)
    print("\n")
items = soup.findAll(id=re.compile("menu"))
print("\nAll elements with id containing 'menu':\n",*items,sep="\n")
articles = soup.find all("article")
print("\nAll article tags:\n",*articles,sep="\n")
```

a)

# b)

```
<div class="theme-title"><span class="theme">TEMAER</span><span class="corner"><svg enable-background="new 0 0 16 32" height="32px" id="Lay
er_1" version="1.1" viewbox="0 0 16 32" width="16px" x="0px" xml:space="preserve" xmlns="http://www.w3.org/2000/svg" xmlns:xlink="http://ww</pre>
er_1" version="1.1" viewbox="0
w.w3.org/1999/xlink" y="0px">
<g>>
</g>
</svg></span></div>

<1i>>
carric lass="views-field views-field-name"> <span class="field-content"><a href="/emne/nye-lufthavne">nye lufthavne</a></span> </div> 
<div class="views-field views-field-name"> <span class="field-content"><a href="/emne/vaccinesagen">Vaccinesagen</a></span> </div> 
<1i>>
div class="views-field views-field-name"> <span class="field-content"><a href="/emne/kuannersuitkvanefjeld">Kuannersuit/Kvanefjeld</a></sp
an> </div> 
<div class="views-field views-field-name"> <span class="field-content"><a href="/emne/sexchikane">Sexchikane</a></span> </div> 

<div class="clearfix"></div> </div>
<div id="omenu"></div>
```

## c)

```
All article tags:
<article class="node node--article node--w615 node--article--w615" role="article">
<h1>Arctic people concerned for the future </h1>
</article>
<article class="node node--article node--w300 node--article--w300" role="article">
<h1>Press release: Common Arctic search and rescue service agreed</h1>
</article>
<article class="node node--article node--w300 node--article--w300" role="article">
<h1>Address Kuupik Kleist Arctic Council 7 Meeting 2011</h1>
</article>
<article class="node node--article node--w300 node--article--w300" role="article">
\hdots (h1>Open letter sent to the Foreign Ministers of Canada, U.S., Norway, Denmark, Greenland and Russia</h1>
<article class="node node--article node--w300 node--article--w300" role="article">
<h1>Resource Development Principles in Inuit Nunaat</h1>
</article>
 <article class="node node--article node--w300 node--article--w300" role="article">
<h1>Arctic Council Nuuk Ministerial Agenda</h1>
<article class="node node--article node--w300 node--article--w300" role="article">
<h1>Editorial: Our friends in Norway</h1>
</article>
<article class="node node--article node--term-list node--article--term-list" role="article">
<h1>Climate change - it's about the people</h1>
  Living conditions and economic development in the face of climate change are the challenges the Arctic Council needs to deal with, Greenl
</article>
<article class="node node--article node--term-list node--article--term-list" role="article">
<h1>Tight security during Arctic summit</h1>
  Terrorism threat levels remain "serious" as Arctic foreign ministers gather in Nuuk</h2>
<article class="node node--article node--term-list node--article--term-list" role="article">
<h1>New Arctic strategy to be presented in Nuuk</h1>
```

```
As climate scientists paint an increasingly dire picture of global warming, it has begun to sink in with lawmakers that slowing the devel opment requires immediate action</h2>
</article>
<article class="node node--article node--term-list node--article--term-list" role="article">
<h1>Antricle Council Nuuk Ministerial Agenda</h1>
<h2>
</h2>
</h2>
</article>
<article class="node node--article node--term-list node--article--term-list" role="article">
<h2>
</h2>
</h2>
</article class="node node--article node--term-list node--article--term-list" role="article">
<h1>Participants in Arctic Council Nuuk Ministerial</h1>
<h2>
The smallest delegations have as few as six members, the largest up to 25</h2>
</article>
```

```
Enter the url: https://sermitsiaq.ag/english
  -> head
    -> link
    -> link
    -> link
    -> script
    -> link
      -> link
        -> meta
        -> script
        -> link
          -> meta
          -> link
          -> link
          -> link
          -> link
          -> meta
          -> meta
          -> link
          -> link
          -> meta
          -> title
          -> link
          -> link
          -> link
          -> style
          -> script
          -> script
          -> script
```

```
-> script
        -> script
        -> script
        -> script
        -> meta
        -> script
        -> script
        -> link
        -> link
        -> link
        -> script
        -> script
        -> script
        -> script
        -> script
-> body
 -> noscript
  | -> iframe
  -> div
  -> script
  -> script
  -> a
  -> header
    -> div
      -> div
        -> div
          -> ul
            -> li
               -> a
               | -> img
               -> a
               | -> img
          -> button
          -> button
```

| -> div   |
|--|
|  |
|  |
|  |
| -> div<br>            -> gcse:search<br>          -> div<br>      -> div<br>      -> div<br>        -> a<br>          -> img<br>      -> div<br>        -> a<br>            -> a<br>              -> a<br>              -> a<br>              -> a<br>              -> a<br>              -> a<br>              -> a<br>              -> a<br>              -> a<br>              -> a<br>              -> a |
|  |
| -> nav<br>            -> div<br>              -> ul<br>                -> a<br>                -> a<br>                  -> a<br>                  -> li<br>                  -> li<br>                  -> li   |
|  |
|  |
|  |
| -> 1i  |
|  |
| -> 1i  |
| -> a<br>                -> li<br>                  -> a<br>                -> li   |
| -> 1i  |
|  |
| -> a<br>                -> li  |
|  |
| -> 1i  |
|  |
| -> 1i  |
|  |
| -> li<br>                -> a<br>                  -> a<br>                  -> a<br>                  -> a<br>                  -> a<br>                  -> a  |
| -> a   |
| -> a<br>                -> li  |
| -> a   |
| -> a<br>                -> li  |
| -> a   |
| -> a<br>                -> li  |
| -> a                 -> li                   -> a                   -> a                   -> a                   -> a                     -> a                     -> a                     -> a                     -> a                       -> a  |
|  |
|  |

| -> p                 -> a           -> div           -> h2             -> h2             -> h2             -> h2             -> a             -> h2               -> a             -> a               -> a               -> a               -> a               -> a               -> a               -> a               -> a               -> a               -> a               -> a               -> b2                 -> a                 -> a                 -> a                 -> div |  |
|---|--|
|   |  |
| -> div<br>            -> h2<br>            -> h2<br>            -> a<br>               a<br>            -> a<br>            -> a<br>            -> a<br>            -> b2<br>              -> b2<br>              -> b2<br>              -> b2  |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
|   |  |
| / """   |  |
|   |  |
|   |  |
|   |  |
| -> script   |  |
| -> script   |  |
|   |  |
| -> p<br>            -> strong<br>    -> script<br>    -> script<br>    -> script<br>    -> script<br>    -> script<br>    -> script<br>    -> script  |  |
| -> script   |  |
| -> script   |  |
| -> script   |  |

- 4. Take website: https://www.batimes.com.ar
  - a. List items of class "nav-item text-uppercase px-0"
  - b. Search for string "Matías Lammens" and list out the HTML item of which the string occurs
  - c. List all the images in the page
  - d. List out the DOM hierarchy of the page.

```
import requests
from bs4 import BeautifulSoup as bs
import re
url=input('Enter the url: ')
resp = requests.get(url)
page = resp.text
soup= bs(page,'html.parser')
items = soup.find_all(class_="nav-item text-uppercase px-0")
print('\nList items of class "nav-item text-uppercase px-
0": ',*items,sep='\n')
elements = soup.find_all(string=re.compile('Matías Kulfas'))
print("\nElements containing string 'Matías Kulfas'")
for elem in elements:
    print(elem.parent)
images = soup.find_all("img")
print("\nAll Images :",*images,sep="\n\n")
```

## a)

```
(vernv) PS C:\Usera\Gencsh Shibu George\Desktop\VIT College\Third Year\Web Minning\Programs> python exercise4.py
Enter the unl: https://www.batimes.com.ar

List items of class "nav-item text-uppercase px-0":
(li class="nav-item text-uppercase px-0":
(a class="px-0 nav-link" href="/last-news/" target="_self" title="Topics of the day">
Topics
(/a)
(/ii)
(li class="nav-item text-uppercase px-0">
(a class="px-0 nav-link" href="/last-news/" target="_self" title="Topics of the day">
Topics
(/a)
(/ii)
(li class="nav-item text-uppercase px-0">
(a class="px-0 nav-link" href="/news/argentina/president-fernandez-under-fire-for-olivos-birthday-party-during-lockdown.phtml" target="_self" title="President Fernandez under fire for Olivos birthday party during lockdown">
Olivos party
(/a)
(/ii)
(li class="nav-item text-uppercase px-0">
(a class="px-0 nav-link" href="/news/sports/messis-hunger-for-more-means-star-is-sure-to-shine-in-city-of-lights.phtml" target="_self" title="Messi's hunger for more means star is sure to shine in City clights">
Messi moves to Paris
(/a)
(/li)
(slass="nav-item text-uppercase px-0")
(a class="px-0 nav-link" href="/news/argentina/matias-kulfas-to-lower-inflation-you-need-multi-sector-agreements-and-clear-objectives.phtml" target="_self" title="Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you need multi-sector agreements and clear objectives/">
Matias Kulfas: To lower inflation you ne
```

# b)

```
Elements containing string 'Matías Kulfas'

(a class="px-0 nav-link text-secondary" href="/news/argentina/matias-kulfas-to-lower-inflation-you-need-multi-sector-agreements-and-clear-objectives.phtml" target = "_self" title="Matías Kulfas: (To lower inflation you need multi-sector agreements and clear objectives'")

Matías Kulfas interview

(/a)

(a class="px-0 nav-link" href="/news/argentina/matias-kulfas-to-lower-inflation-you-need-multi-sector-agreements-and-clear-objectives.phtml" target="_self" title=
"Matías Kulfas: (To lower inflation you need multi-sector agreements and clear objectives'")

Matías Kulfas: interview

(/a)

(h2)

(h2)

(h2)

(h2)

(h3)

(h2)

(h2)

(h2)

(h2)

(h3)

(h2)

(h4)

(5pan)13-08-2021 22:36(/span) Productive Development Minister Matías Kulfas on trying to lower inflation, the IMF agreement and the role of the state.

(/h4)
```

# c)

<img alt="when weakness is strength" class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/when-weakness-is-strength-1223032.jpg 270w,https://fotos.perfil.com/2021/08/27/trim/720/405/when-weakness-is-strength-1223032.jpg 540w,https://fotos.perfil.com/2021/08/27/trim/720/405/when-weakness-is-strength-1223032.jpg 720w,https://fotos.perfil.com/2021/08/27/trim/720/355/when-weakness-is-strength-1223032.jpg 720w," src="https://via.placeholder.com/720x405?text=BATIMES"/>

<img alt="In Buenos Aires, it's known that many of these game-changing businessmen were trained at the Instituto Tecnológico de Buenos Aires (ITBA), a great devel
oper of local talent. " class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/instituto-tecnologico-de-buenosaires-itba-1222983.jpg 270w,https://fotos.perfil.com/2021/08/27/trim/540/304/instituto-tecnologico-de-buenos-aires-itba-1222983.jpg 720w,https://fotos.perfil.com/2021/08/27/trim/720/405/instituto-tecnologico-de-buenos-aires-itba-1222983.jpg 720w," src="https://via.placeholder.com/720x405?text=BATIMES"/>

<img alt="Cristiano Ronaldo." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/cristiano-ronaldo-1222742.jpg 270w, https://fotos.perfil.com/2021/08/27/trim/720/405/cristiano-ronaldo-1222742.jpg 540w, https://fotos.perfil.com/2021/08/27/trim/720/405/cristiano-ronaldo-1222742.jpg 720w, https://fotos.perfil.com/2021/08/27/trim/720/355/cristiano-ronaldo-1222742.jpg 720w," src="https://via.placeholder.com/720x405?text=BATIMES"/>

<img alt="Malcolm Hunter OBE." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/malcolm-hunter-obe-1223024.jpg 270w,https://fotos.perfil.com/2021/08/27/trim/720/405/malcolm-hunter-obe-1223024.jpg 540w,https://fotos.perfil.com/2021/08/27/trim/720/405/malcolm-hunter-obe-1223024.jpg 720w,https://fotos.perfil.com/2021/08/27/trim/720/355/malcolm-hunter-obe-1223024.jpg 720w," src="https://via.placeholder.com/720x405?text=BATIMES"/>

cimg alt="The image that rocked the week - a secret clandestine birthday party for First Lady Fabiola Yáñez took place at the Olivos presidential residence last J
une, during Argentina's strict coronavirus lockdown." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/14/trim/270/152/ol
ivos-birthday-lockdown-party-1217754.jpg 270w,https://fotos.perfil.com/2021/08/14/trim/540/304/olivos-birthday-lockdown-party-1217754.jpg 540w,https://fotos.perfil.com/2021/08/14/trim/720/355/olivos-birthday-lockdown-party-1217754.jpg 720w,https://fotos.perfil.com/2021/08/14/trim/720/355/olivos-birthday-lockdown-party-1217754.jpg 720w,https://fotos.perfil.com/2021/08/14/trim/720/355/olivos-birthday-lockdown-party-12177
754.jpg 720w," src="https://via.placeholder.com/720x405?text=8ATIMES"/>

cimg alt="Members of Spanish National Police forces and military personnel who were stationed in Afghanistan disembark from one of the last two Spanish evacuation
flights, at the Torrejon de Ardoz air base, 30 km from Madrid, on August 27, 2021. Spain announced the end of its Kabul evacuations today after a nine-day operat
ion that saw more than 2,200 people flown out of the strife-torn country following the Taliban takeover. " class='img-fluid lazyload" data-sizes="720" data-srcset
="https://fotos.perfil.com/2021/08/27/trim/270/152/kabul-airlift-1222792.jpg 270w,https://fotos.perfil.com/2021/08/27/trim/750/405/kabul-airlift-1222792.jpg 720w,https://fotos.perfil.com/2021/08/27/trim/720/355/kabul-airlift-1222792.jpg 720w,"
src="https://via.placeholder.com/720x405?text=BATIMES"/>

<img alt="US President Joe Biden gives a press conference on the situation in Afghanistan." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/us-president-joe-biden-1222548.jpg 270w,https://fotos.perfil.com/2021/08/27/trim/540/304/us-president-joe-biden-1222548.jpg 54
0w,https://fotos.perfil.com/2021/08/27/trim/720/405/us-president-joe-biden-1222548.jpg 720w,https://fotos.perfil.com/2021/08/27/trim/720/355/us-president-joe-biden-1222548.jpg</pre>

/270/152/passengers-ezeiza-airport-tourism-coronavirus-covid-1221313.jpg 270w, https://fotos.perfil.com/2021/08/24/trim/540/304/passengers-ezeiza-airport-tourism-coronavirus-covid-1221313.jpg 540w, https://fotos.perfil.com/2021/08/24/trim/720/405/passengers-ezeiza-airport-tourism-coronavirus-covid-1221313.jpg 720w, https://fotos.perfil.com/2021/08/24/trim/720/405/passengers-ezeiza-airport-tourism-coronavirus-covid-1221313.jpg 720w," src="https://via.placeholder.com/720x405?text=BATIME 57"/>

<img alt="Mohsen Rezai." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/26/trim/270/152/mohsen-rezai-1222527.jpeg 270w,
https://fotos.perfil.com/2021/08/26/trim/540/304/mohsen-rezai-1222527.jpeg 480w,https://fotos.perfil.com/2021/08/26/trim/240/85/mohsen-rezai-1222527.jpeg 720w, https://fotos.perfil.com/2024/08/26/trim/720/355/mohsen-rezai-1222527.jpeg 720w,"britps://fotos.perfil.com/2024/08/26/trim/720/355/mohsen-rezai-1222527.jpeg 720w,"britps://fotos.perfil.com/2024/08/26/trim/720/352/mohsen-rezai-1222527.jpeg 720

<img alt="In latin America and the Caribbean, more than 27 percent of the population live in coastal areas." class="img-fluid lazyload" data-sizes="720" data-srcs
et="https://fotos.perfil.com/2021/08/25/trim/270/152/climate-change-rising-sea-levels-latin-america-1221727.jpeg 270w,https://fotos.perfil.com/2021/08/25/trim/740/20405/climate-change-rising-sea-levels-latin-america-1221727.jpeg 370w,https://fotos.perfil.com/2021/08/25/trim/720/405/climate-change-rising-sea-levels-latin-america-1221727.jpeg 720w,https://fotos.perfil.com/2021/08/25/trim/720/405/climate-change-rising-sea-levels-latin-america-1221727.jpeg 720w," src="https://via.placeholder.com/720x405?text=BATIMES"/>

cimg alt="Atucha Nuclear Complex in Lima, Zárate, Buenos Aires Province." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08
8/21/trim/270/152/atucha-nuclear-complex-in-lima-zarate-buenos-aires-province-1220486.jpg 270w,https://fotos.perfil.com/2021/08/21/trim/20/405/atucha-nuclear-complex-in-lima-zarate-buenos-aires-province-1220486.jpg 700w,https://fotos.perfil.com/2021/08/21/trim/720/405/atucha-nuclear-complex-in-lima-zarate-buenos-aires-province-1220486.jpg 700w,https://fotos.perfil.com/2021/08/21/trim/720/355/atucha-nuclear-complex-in-lima-zarate-buenos-aires-province-1220486.jpg 720w," src="https://via.placeholder.com/720x405?text=BATIMES"/>

<img alt="Productive Development Minister Matias Kulfas." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/13/trim/270/15
2/matias-kulfas-1217683.jpg 270w,https://fotos.perfil.com/2021/08/13/trim/540/304/matias-kulfas-1217683.jpg 720w,https://fotos.perfil.com/2021/08/13/trim/720/355/matias-kulfas-1217683.jpg 720w," src="https://via.placeholder.com/720x405?text=BAT
IMES"/>

<img alt="Former Uruguay president Julio María Sanguinetti." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270
/152/julio-maria-sanguinetti-1222994.jpg 270w,https://fotos.perfil.com/2021/08/27/trim/540/304/julio-maria-sanguinetti-1222994.jpg 540w,https://fotos.perfil.com/2021/08/27/trim/720/405/julio-maria-sanguinetti-1222994.jpg 720w,https://fotos.perfil.com/2021/08/27/trim/720/355/julio-maria-sanguinetti-1222994.jpg 720w," src="https://ia.placeholder.com/720x4057text=BAITMES"/>

(img alt="Maria Eugenia Vidal, pictured during an interview with Perfil's Jorge Fontevecchia." class="img-fluid lazyload" data-sizes="720" data-srcset="https://fo
tos.perfil.com/2021/08/07/trim/270/152/maria-eugenia-vidal-1214859.jpg 270w,https://fotos.perfil.com/2021/08/07/trim/540/304/maria-eugenia-vidal-1214859.jpg 540w,
https://fotos.perfil.com/2021/08/07/trim/720/355/maria-eugenia-vidal-1214859.jpg 720w,https://fotos.perfil.com/2021/08/07/trim/720/355/maria-eugenia-vidal-1214859
jpg 720w," src="https://via.placeholder.com/720x405?text=BATIMES"/)

<img alt="Luis D'Elia celebrates." class="img-fluid lazyload" data-sizes="1140" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/luis-delia-1222992.jp
g 270w,https://fotos.perfil.com/2021/08/27/trim/720/352/luis-delia-1222992.jpg 540w,https://fotos.perfil.com/2021/08/27/trim/720/405/luis-delia-1222992.jpg 720w,
https://fotos.perfil.com/2021/08/27/trim/960/540/luis-delia-1222992.jpg 720w,
https://fotos.perfil.com/2021/08/27/trim/960/540/luis-delia-1222992.jpg 960w,https://
fotos.perfil.com/2021/08/27/trim/1140/540/luis-delia-1222992.jpg 1140w,
https://fotos.perfil.com/2021/08/27/trim/1140/540/luis-delia-1222992.jpg 1140w,
https://fotos.perfil.com/2021/08/27/trim/1140/540/luis-delia-1222992.jpg 1140w,
https://dia.placeholder.com/1140x540/tuis-delia-1222992.jpg 1140w,
https://dia.placeholder.com/1140x540/tuis-deli

<img alt="Bullish about the economy, or just bull?" class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/bull
ish-about-the-economy-or-just-bull-1222987.jpg 270w,https://fotos.perfil.com/2021/08/27/trim/540/304/bullish-about-the-economy-or-just-bull-1222987.jpg 540w,https:
//fotos.perfil.com/20/21/08/27/trim/720/405/bullish-about-the-economy-or-just-bull-1222987.jpg 720w," src="https://via.placeholder.com/720x355?text=BATIMES"/>

<img alt="In this file photo taken on February 15, 2016, The Rolling Stones (left to right) - Mick Jagger, Charlie Watts, Keith Richards and Ron Wood - are pictur
ed upon landing in Montevideo. Charlie Watts, drummer with legendary British rock'n'roll band the Rolling Stones, died on August 24, 2021 aged 80, according to a
statement from his publicist. " class="img-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/24/trim/270/152/rolling-stones-1221290.j
gp 270w,https://fotos.perfil.com/2021/08/24/trim/720/485/rolling-stones-1221290.j
gp 720w,https://fotos.perfil.com/2021/08/24/trim/720/355/rolling-stones-1221290.j
gp 720w,https://fotos.perfil.com/2021/08/24/trim/720/485/rolling-stones-1221290.j
g alt="Bullish about the economy, or just bull?" class="img-fluid lazyload" data-sizes="270" data-srcset="https://fotos.perfil.com/2021/08/27/trim/270/152/bull
ish-about-the-economy-or-just-bull-1222987.jpg 270w," src="https://via.placeholder.com/270x152?text=BATIMES"/>

<img alt="Olivos-Gate: Secret birthday parties." class="img-fluid lazyload" data-sizes="270" data-srcset="https://fotos.perfil.com/2021/08/21/trim/270/152/olivos-gate-secret-birthday-parties-1220360.jpg 270w," src="https://via.placeholder.com/270x152?text=BATIMES"/>

<img alt="Is moderation in pursuit of justice no longer a virtue?" class="img-fluid lazyload" data-sizes="270" data-srcset="https://fotos.perfil.com/2021/08/14/tr
im/270/152/editorial-art-1217752.jpg 270w," src="https://via.placeholder.com/270x152?text=BATIMES"/>

<img alt="Olympian effort?" class="img-fluid lazyload" data-sizes="270" data-srcset="https://fotos.perfil.com/2021/08/07/trim/270/152/alberto-olympics-art-1214857
.jpg 270w," src="https://via.placeholder.com/270x152?text=BATIMES"/>

<img alt="PASO games" class="img-fluid lazyload" data-sizes="270" data-srcset="https://fotos.perfil.com/2021/07/31/trim/270/152/paso-games-1211615.jpg 270w," src=
"https://via.placeholder.com/270x152/text=BATIMES"/>

<img alt="Workers bury a victim of Covid-19 at the Chacarita cemetery in Buenos Aires." class="img-fluid lazyload" data-sizes="270" data-srcset="https://fotos.per fil.com/2021/07/17/trim/270/152/covid-chacarita-cemetery-buenos-aires-covid-1205664.jpg 270w," src="https://via.placeholder.com/270x152?text=BATIMES"/> 04/indigenous-protest-brasilia-brazil-1221099.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/355/indigenous-protest-brasilia-brazil-1221099.jpg 720w," src="https://via.placeholder.com/720x4057text-8ATIMES"/>

Cing alt="In this file photo taken on July 22, 2021 soldiers stand guard outside the Sierra Centro Norte prison in Latacunga, Ecuador, after a riot which left 21 dead occurred." class="ing-fluid lazyload" data-sizes="720" data-srcset="https://fotos.perfil.com/2021/08/23/trim/7210/152/ecuador-prison-stock-1221098.jpg 720w,https://fotos.perfil.com/2021/08/23/trim/720/08/25/ecuador-prison-stock-1221098.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/08/25/ecuador-prison-stock-1221098.jpg 720w," src="https://fotos.perfil.com/2021/08/23/trim/720/05/ecuador-prison-stock-1221098.jpg 720w," src="https://fotos.perfil.com/2021/08/23/trim/720/05/ecuador-prison-stock-1221098.jpg 720w," src="https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-miniater-mart-acratabia-12200995.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-miniater-mart-acratabia-12200995.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-12200995.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-12200995.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-12200995.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-12200995.jpg 540w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-1220095.jpg 720w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-1220095.jpg 720w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-1220095.jpg 720w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-acratabia-1220095.jpg 720w,https://fotos.perfil.com/2021/08/23/trim/720/05/jtalian-justice-minister-mart-ac

<img src="https://b.scorecardresearch.com/p?c1=2&amp;c2=6906401&amp;cv=2.0&amp;cj=1">
</img>
Total number of image tags: 94
(venv) PS C:\Users\Gerosh Shibu George\Desktop\VIT College\Third Year\Web Minning\Programs>

### d)

(venv) PS C:\Users\Gerosh Shibu George\Desktop\VIT College\Third Year\Web Minning\Programs> python find\_dom.py Enter the url: https://www.batimes.com.ar -> head -> meta -> title -> link -> link -> link -> link -> link -> link -> link -> link -> link

| -> link<br>            -> link<br>            -> script<br>            -> script<br>            -> script<br>            -> script  |
|---|
|   |
|   |
| -> link<br>            -> script<br>            -> script<br>            -> script<br>            -> script<br>  -> body<br>  -> script<br>  -> noscript<br>    -> iframe<br>    -> div<br>      -> div<br>        -> div<br>          -> div<br>          -> div<br>            -> a<br>            -> img<br>        -> div<br>            -> nav |
|   |
| -> script   |
| -> script   |
| -> noscript   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
|   |
| -> div  |
| -> span<br>              -> img<br>                -> div<br>                -> ul<br>                  -> li<br>                  -> li<br>                  -> li   |
|   |
|   |
|   |
| -> 1i   |
|   |
|   |
|   |
|   |

| $\Box$ | -> div     |
|--------|------------|
| i i    |            |
| i i    |            |
|        |            |
|        | -> li      |
|        | -/ 11      |
|        |            |
|        |            |
| !!     | -> a       |
| !!     |            |
| !!     | -> a       |
|        |            |
|        | -> a       |
|        | -> div     |
|        |            |
|        | ->i        |
|        |            |
|        | -> i       |
|        |            |
|        | -> i       |
|        |            |
| Ιİ     |            |
|        | -> main    |
|        | -> section |
|        | -> section |
|        |            |
| Ιİ     |            |
| i i    |            |
| i i    |            |
| i i    |            |
|        |            |
| i i    | -> div     |
|        |            |
|        | -> div     |
|        |            |
|        |            |
|        |            |
| 1 1    | -> a       |

|  | а                                |
|--|----------------------------------|
|  | ·> figure                        |
|  | -> figure<br>-> div              |
|  | iv                               |
|  | div                              |
|  | ·> article                       |
|  | div<br>-> article<br>-> article  |
|  | iv                               |
| ->,  | article  > a article article > a |
|  | ·> a                             |
| ->,  | article                          |
|  | ·> a                             |
| -><br>                        -><br>                          -><br>                          -><br>                          -><br>                      -> div | iv                               |
| ->   | div                              |
|  | iv                               |
| ->   | div                              |
|  |                                  |
|  | iv                               |
|  |                                  |
| -> div   |                                  |
|  |                                  |
|  | le                               |
|  |                                  |
|  | igure                            |
| ->   | img                              |
|  | iv                               |
| ->   | h3                               |
| ->   | h2                               |
| ->   | h4                               |
| ->   | h5                               |
| -> div   |                                  |
| -> div   |                                  |
|  | rticle                           |
| ->   | а                                |
|  | ·> figure                        |
| <u> </u>   | -> div                           |

| 1  | 1 1 |    | - 1 | 1  | 1   |            |
|----|-----|----|-----|----|-----|------------|
| -  | 1 1 |    | - 1 | -  | 1   |            |
|    | IJ  |    | - [ |    | 1   | -> div     |
|    |     |    |     |    |     |            |
|    |     |    |     | Т  |     | -> article |
| ĺ  | İΪ  | Ιİ | İ   | İ  | İ   |            |
| i  | Ιi  | i  | i   | i  | i   |            |
| i  | ii  | i  | i   | i  | i   |            |
| i  | ii  | H  | i   | i  | i   |            |
| -  | 1 1 |    | H   | i. | 1   |            |
| -  |     |    | -   | -  | -   |            |
| !  | ! ! |    | -   | !  | !   |            |
| !  | !!  |    | !   | ļ  | !   |            |
| İ  |     |    | ļ   | İ  | ļ   |            |
| Ţ  | ļΙ  |    |     | Ţ  | Ţ   | -> article |
|    |     |    |     |    |     | -> span    |
|    |     |    |     |    |     |            |
|    |     |    |     |    |     |            |
| Ĺ  | İΪ  |    | Ĺ   | Ĺ  | Ĺ   |            |
| i  | i i | i  | i   | i  | i   |            |
| i  | ii  | i  | i   | i  | i   |            |
| i  | ii  | i  | i   | i  | i   |            |
| i  | 1 1 |    | H   | i  | i i |            |
| -  |     |    | -   | -  | 1   | -> div     |
| -  |     |    | -   | -  | -   | : .        |
| !  | ! ! |    | -   | !  | !   |            |
| !  | !!  |    | !   | !  | !   | -> article |
| ļ. | ļ ļ |    | Ţ   | ļ  | ļ.  | -> span    |
|    | ļΙ  |    |     |    |     |            |
|    |     |    |     |    |     |            |
|    |     | ĺĺ |     |    |     |            |
| i  | j   | i  | i   | į  | İ   |            |
| i  |     |    | i   | i  | i   |            |
|    |     |    |     |    |     |            |
| -  |     |    |     |    |     |            |
|    |     |    | ļ   |    |     |            |
|    |     |    | ļ   | ļ  | ļ   |            |
|    |     |    |     |    |     | -> article |
|    |     |    |     |    |     | -> span    |
|    |     | ĺ  | İ   |    |     |            |
|    |     |    |     |    |     |            |

| -> article<br>                -> span<br>                  -> a<br>                    -> figure<br>                      -> img<br>                    -> h4<br>                  -> h3<br>                  -> h2<br>      -> div<br>          -> div<br>            -> img<br>                -> img<br>                -> img<br>                -> img<br>                -> img<br>                -> img<br>                -> img<br>                  -> img |
|---|
|   |
|   |
| -> div  |
|   |
| -> a  |
|   |
|   |
|   |
| -> img  |
| -> div  |
|   |
|   |
| -> i  |
|   |
| -> i  |
|   |
| -> i<br>              -> a<br>                -> i<br>                -> i<br>                -> i<br>                -> a  |
| -> a  |
| -> i  |
|   |
| -> i  |
| -> i<br>              -> a<br>              -> i<br>              -> i<br>                -> i<br>                -> i  |
| -> i  |
| -> a<br>              -> i<br>              -> a<br>              -> a<br>              -> a<br>              -> i<br>          -> nav  |
| -> i<br>              -> a<br>              -> i<br>              -> i<br>              -> a<br>                -> i<br>              -> i<br>              -> ul   |
| -> ul   |

```
-> a
             -> li
              -> a
      -> div
        -> div
            -> span
            -> br
             -> br
             -> div
               -> span
               -> span
               -> span
               -> span
               -> span
               -> a
                -> span
                 | -> spa
  -> div
    -> button
    | -> i
    -> div
  -> div
-> script
-> script
-> script
-> script
-> script
-> script
-> script
-> script
-> script
-> script
-> noscript
 -> img
```

# CODE FOR FINDING THE DOM STRUCTURE:

```
import requests
from bs4 import BeautifulSoup as bs
url=input('Enter the url: ')
resp = requests.get(url,headers={'User-Agent': 'Mozilla/5.0'})
page = resp.text
soup= bs(page,'html.parser')
def graph(value, size):
    if size>=12:
        return
    else:
        for i in value.findChildren(recursive=False):
            for j in range(-1,size,1):
                print("|",end=" ")
            print("->",i.name)
            graph(i,size+1)
graph(soup.html,0)
```