WEB MINING
LAB
CSE 3024

Faculty: Dr.Sridhar.R

LAB1
DATE: 5TH AUG 2021
VIT CHENNAI

Aim : To understand request structure of a web app

Q1. Request structure

Encoding:

The Encoding API provides a mechanism for handling text in various character encodings, including legacy non-UTF-8 encodings. The API provides four interfaces: TextDecoder , TextEncoder , TextDecoder Stream and TextEncoderStream .

Code:

```
import requests
 URL = "https://chennai.vit.ac.in/"
response = requests.get(URL,verify=False)


print("\n\nEncoding-" + response.encoding)
```

OutPut:

```
Encoding-UTF-8
apple@Apples-MacBook-Pro WebMining %
```

Status code:

HTTP response status codes indicate whether a specific HTTP request has been successfully completed.

Code:

```
import requests
 URL = "https://chennai.vit.ac.in/"
response = requests.get(URL,verify=False)


print("\n\nStatus code-" + response.status_code)
```

OutPut:

```
Status code-200
apple@Apples-MacBook-Pro WebMining %
```

Headers:

A request header is an HTTP header that can be used in an HTTP request to provide information about the request context, so that the server can tailor the response

Code:

```
import requests
import json
URL = "https://chennai.vit.ac.in/"
response = requests.get(URL,verify=False)

print(json.dumps(response.json, indent = 1))
```

OutPut:

```
Headers{'Date': 'Sun, 08 Aug 2021 15:25:33 GMT', 'Server': 'Apache', '
X-Frame-Options': 'SAMEORIGIN', 'Vary': 'Accept-Encoding', 'X-Powered-
By': 'PHP/7.2.27', 'Last-Modified': 'Sun, 08 Aug 2021 12:30:21 GMT', '
Content-Encoding': 'gzip', 'Content-Length': '36786', 'Keep-Alive': 't
imeout=5, max=100', 'Connection': 'Keep-Alive', 'Content-Type': 'text/
html; charset=UTF-8'}
```

History

Past responses can be viewed in the dropdown menu of the response pane.

Code:

```
import requests
URL = "https://chennai.vit.ac.in/"
response = requests.get(URL,verify=False)

print("\n\nHistory")
print(response.history)
```

OutPut:

```
History
[]
```

Cookies:

Cookies are created to identify you when you visit a new website.

```
import requests
URL = "https://chennai.vit.ac.in/"
response = requests.get(URL,verify=False)

print("\n\nCookies")
print(response.cookies)
```

OutPut:

```
Cookies
<RequestsCookieJar[]>
```

Elapsed:

response.elapsed returns a timedelta object with the time elapsed from sending the request to the arrival of the response.

```
import requests
URL = "https://chennai.vit.ac.in/"
response = requests.get(URL,verify=False)

print("\n\nElapsed")
print(response.elapsed)
```

OutPut:

```
Elapsed
0:00:00.095294
```

Q1 .With the help of the ‘Requests’ module to download and analyze page contents of ‘https://www.vit.ac.in/academics/home ‘ or ‘https://www.vit.ac.in/admissions/overview’ or any other page in these menu tabs.

Code:

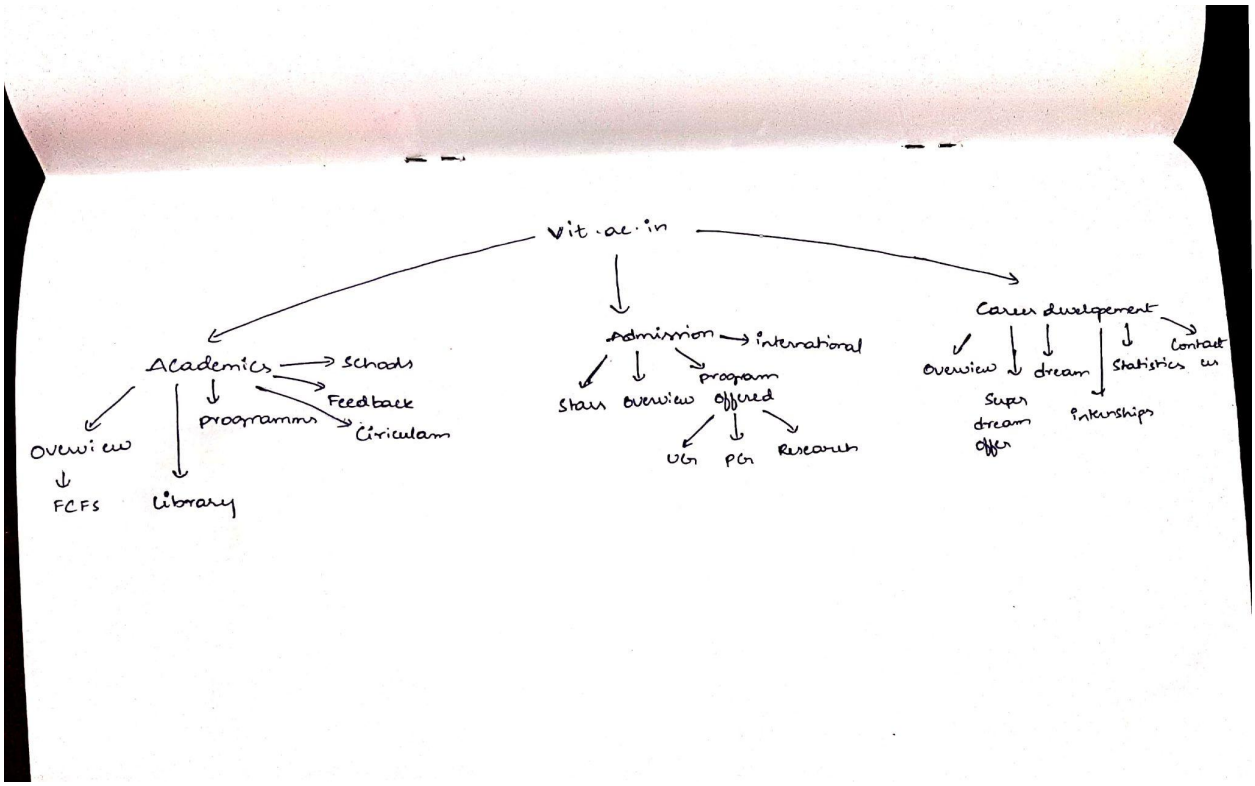
```
import requests
from bs4 import BeautifulSoup

URL = "https://www.vit.ac.in/admissions/overview’"
response = requests.get(URL,verify=False)

soup = BeautifulSoup(response.text, 'lxml')

print(soup.prettify())
```

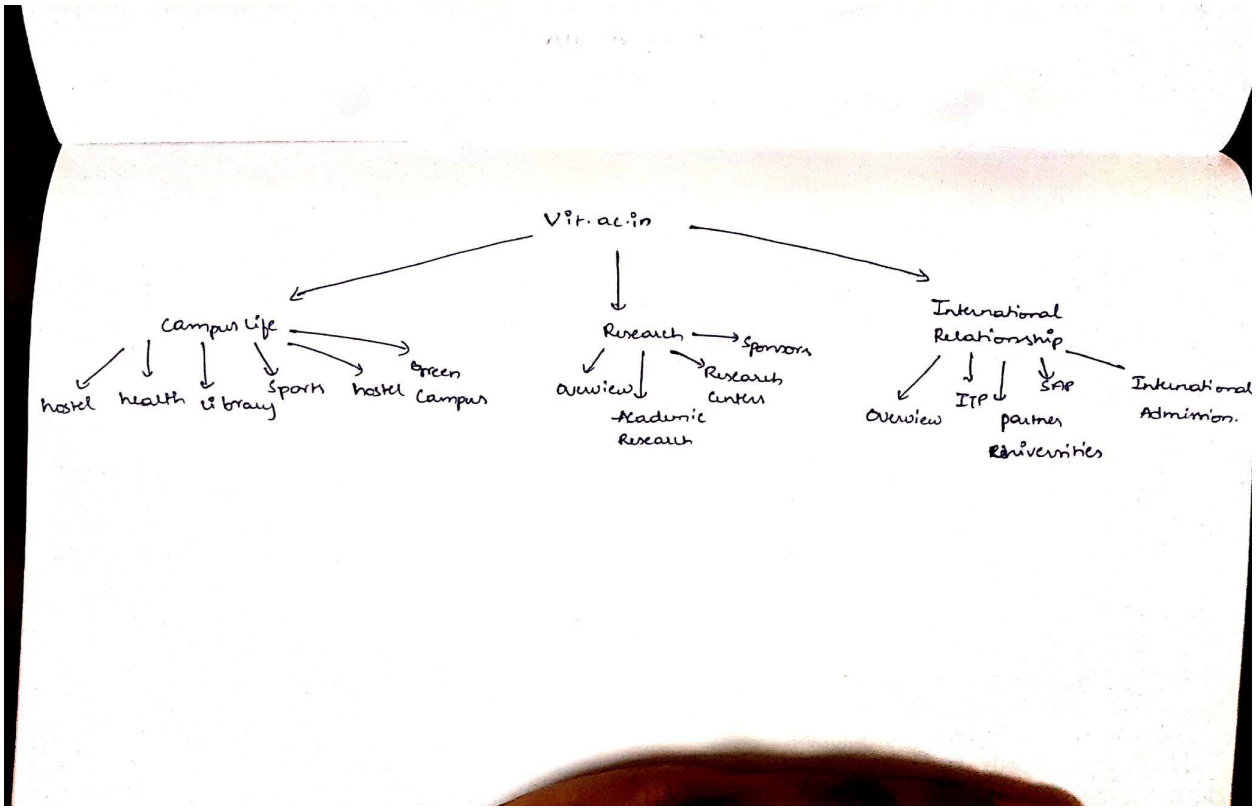
OutPut:

Below is the content of the page

```
    <header class="clearfix section inner_cnt">
     <div class="container">
      <div class="ban_qs_img">
       <img alt="VIT-QS" src="/sites/all/themes/vittheme/images/ban_star.jpg" title="VIT-QS"/>
      </div>
      <div class="identity-stripe">
      </div>
      <div class="banner_block">
       <div class="section inner_banner">
        <!-- ASC Slider -->
        <div class="region region-customheader">
         <div class="block block-headerimage clearfix" id="block-headerimage-2">
          <div class="content">
           <div class="field field-name-field-header-banner-image field-type-image field-label-hidde
            <div class="field-items">
             <div class="field-item even">
              <img alt="UG and PG Admissions in VIT" height="436" src="https://vit.ac.in/sites/defau
anner/014_0.jpg" title="UG and PG Admissions in VIT" width="1349"/>
             </div>
            </div>
           </div>
          </div>
         </div>
        </div>
        <div class="bread_crump center">
         <div class="bread-crumbs">
          <span class="bread_img icons">
          </span>
          <ul class="bread_crump_list">
           <li>
            Admissions
           </li>
          </ul>
         </div>
        </div>
       </div>
      </div>
     </div>
    </header>
    <div class="page_top_fix">
     <!--<a class="main-menu_icon"> <span class="menu-icon"> <span class="menu_txt">Menu</span> <spa
```

Q2. Identify all the sub links in a web app and make a tree of sublinks




Vit. ac.in
Campus Life
hostel
health
Library
Sports
hostel
Green Campus
Research
Sponsors
Research centers
overview
Academic Research
International Relationship
Overview
ITP
partner Universities
SAP
International Admission.