

CSE 3024

Web Mining

LAB ASSESSMENT - 2

NAME: Vibhu Kumar Singh

REG. NO: 19BCE0215

TEACHER: Mr. Hiteshwar Kumar Azad

1. Write a python program to scrape website to extract the following:

- a) raw HTML content
- b) tags (title, p, a, div)
- c) all textual content.

Note: Consider the input to be any website of your choice.

Ans 1.

- a) **Raw HTML content**

HANDWRITTEN CODE:

```
a) VIBHU KUMAR SINGH 19BCE0215

import requests
from bs4 import BeautifulSoup
URL = "https://google.com"

r = requests.get(URL)
soup = BeautifulSoup(r.content, 'html5lib')

print('\nRaw HTML Content:\n')
print(soup.prettify())
```

CODE:

```
import requests
from bs4 import BeautifulSoup
URL = "https://google.com"

r=requests.get(URL)
soup=BeautifulSoup(r.content, 'html5lib')

print('\nRaw HTML Content:\n')
print(soup.prettify())
```

CODE SCREENSHOT:

```
C:\Users\Vibhu\OneDrive - vit.ac.in\Desktop\Winter Semester 21-22\Web Mining\ELA\LAB-2\rawHTML.py
rawHTML.py x
1 import requests
2 from bs4 import BeautifulSoup
3 URL = "https://google.com"
4
5 r=requests.get(URL)
6 soup=BeautifulSoup(r.content, 'html5lib')
7
8 print('\n033[4mRaw HTML Content:\033[0m ')
9 print(soup.prettify())
```

OUTPUT:

Raw HTML Content:

```
<!DOCTYPE html>
<html itemscope="" itemtype="http://schema.org/WebPage" lang="en-IN">
<head>
  <meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>
  <meta content="/images/branding/googleg/1x/googleg_standard_color_128dp.png"
itemprop="image"/>
  <title>
    Google
  </title>
  <script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function){window.google={k...
```

OUTPUT SCREENSHOT:

```
Console 1/A
Python 3.7.6 (default, Jan 8 2020, 20:23:39) [MSC v.1916 64 bit (AMD64)]
Type "copyright", "credits" or "license()" for more information.

IPython 7.12.0 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/Vibhu/OneDrive - vit.ac.in/Desktop/Winter Semester 21-22/Web Mining/ELA/LAB-2/rawHTML.py', wdir='C:/Users/Vibhu/OneDrive - vit.ac.in/Desktop/Winter Semester 21-22/Web Mining/ELA/LAB-2')

Raw HTML Content:
<!DOCTYPE html>
<html itemscope="" itemtype="http://schema.org/WebPage" lang="en-IN">
<head>
  <meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>
  <meta content="/images/branding/googleg/1x/googleg_standard_color_128dp.png" itemprop="image"/>
  <title>
    Google
  </title>
  <script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,7108,113,2146,307,2558,5,37,5,466,945,100,223,416,59,1,2,129,56,2,2260,4144,3956,343,3038,750,38,1524,636,3374,1769,12,2,53,12,286,1839,255,759,1085,229,4,84,640,48,1539,485,660,3,642,852,352,368,859,330,9,496,67,179,370,565,82,401,171,711,702,883,350,95,210,468,410,959,5476301,822,108,5996661,2800420,882,444,1,2,80,1,1796,1,2562,1,748,936,4,559,1,4265,1,1,2,1331,4142,2609,155,17,13,72,139,4,2,20,2,169,13,19,46,5,39,96,548,29,2,2,1,2,1,2,2,7,4,1,2,2,2,2,2,353,513,186,1,1,158,3,2,2,2,2,2,4,2,3,3,269,1601,141,82,4,490,139,1,10,53,45,34,4,38,2,2,38,2,23952036,4041690,3,450,3455,9,1435,159,1358,4726,3,955,3,1215,489,1900,1020,609,1830',kBL:'b0B6'};google.sn='webhp';google.kHL='en-IN'}};})();(function){var f=this||self;var h,k=[];function l(a){for(var b;a&&(l=a.getAttribute)||!(b=a.getAttribute("eid")));)a=a.parentNode;return b||h}function m(a){for(var b=null;a&&(l=a.getAttribute)||!(b=a.getAttribute("leid")));)a=a.parentNode;return b}function n(a,b,c,d,g){var e="";c||!l==b.search("<script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,7108,113,2146,307,2558,5,37,5,466,945,100,223,416,59,1,2,129,56,2,2260,4144,3956,343,3038,750,38,1524,636,3374,1769,12,2,53,12,286,1839,255,759,1085,229,4,84,640,48,1539,485,660,3,642,852,352,368,859,330,9,496,67,179,370,565,82,401,171,711,702,883,350,95,210,468,410,959,5476301,822,108,5996661,2800420,882,444,1,2,80,1,1796,1,2562,1,748,936,4,559,1,4265,1,1,2,1331,4142,2609,155,17,13,72,139,4,2,20,2,169,13,19,46,5,39,96,548,29,2,2,1,2,1,2,2,7,4,1,2,2,2,2,2,353,513,186,1,1,158,3,2,2,2,2,2,4,2,3,3,269,1601,141,82,4,490,139,1,10,53,45,34,4,38,2,2,38,2,23952036,4041690,3,450,3455,9,1435,159,1358,4726,3,955,3,1215,489,1900,1020,609,1830',kBL:'b0B6'};google.sn='webhp';google.kHL='en-IN'}};})();(function){var f=this||self;var h,k=[];function l(a){for(var b;a&&(l=a.getAttribute)||!(b=a.getAttribute("eid")));)a=a.parentNode;return b||h}function m(a){for(var b=null;a&&(l=a.getAttribute)||!(b=a.getAttribute("leid")));)a=a.parentNode;return b}function n(a,b,c,d,g){var e="";c||!l==b.search("<script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,7108,113,2146,307,2558,5,37,5,466,945,100,223,416,59,1,2,129,56,2,2260,4144,3956,343,3038,750,38,1524,636,3374,1769,12,2,53,12,286,1839,255,759,1085,229,4,84,640,48,1539,485,660,3,642,852,352,368,859,330,9,496,67,179,370,565,82,401,171,711,702,883,350,95,210,468,410,959,5476301,822,108,5996661,2800420,882,444,1,2,80,1,1796,1,2562,1,748,936,4,559,1,4265,1,1,2,1331,4142,2609,155,17,13,72,139,4,2,20,2,169,13,19,46,5,39,96,548,29,2,2,1,2,1,2,2,7,4,1,2,2,2,2,2,353,513,186,1,1,158,3,2,2,2,2,2,4,2,3,3,269,1601,141,82,4,490,139,1,10,53,45,34,4,38,2,2,38,2,23952036,4041690,3,450,3455,9,1435,159,1358,4726,3,955,3,1215,489,1900,1020,609,1830',kBL:'b0B6'};google.sn='webhp';google.kHL='en-IN'}};})();(function){var f=this||self;var h,k=[];function l(a){for(var b;a&&(l=a.getAttribute)||!(b=a.getAttribute("eid")));)a=a.parentNode;return b||h}function m(a){for(var b=null;a&&(l=a.getAttribute)||!(b=a.getAttribute("leid")));)a=a.parentNode;return b}function n(a,b,c,d,g){var e="";c||!l==b.search("<script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,7108,113,2146,307,2558,5,37,5,466,945,100,223,416,59,1,2,129,56,2,2260,4144,3956,343,3038,750,38,1524,636,3374,1769,12,2,53,12,286,1839,255,759,1085,229,4,84,640,48,1539,485,660,3,642,852,352,368,859,330,9,496,67,179,370,565,82,401,171,711,702,883,350,95,210,468,410,959,5476301,822,108,5996661,2800420,882,444,1,2,80,1,1796,1,2562,1,748,936,4,559,1,4265,1,1,2,1331,4142,2609,155,17,13,72,139,4,2,20,2,169,13,19,46,5,39,96,548,29,2,2,1,2,1,2,2,7,4,1,2,2,2,2,2,353,513,186,1,1,158,3,2,2,2,2,2,4,2,3,3,269,1601,141,82,4,490,139,1,10,53,45,34,4,38,2,2,38,2,23952036,4041690,3,450,3455,9,1435,159,1358,4726,3,955,3,1215,489,1900,1020,609,1830',kBL:'b0B6'};google.sn='webhp';google.kHL='en-IN'}};})();(function){var f=this||self;var h,k=[];function l(a){for(var b;a&&(l=a.getAttribute)||!(b=a.getAttribute("eid")));)a=a.parentNode;return b||h}function m(a){for(var b=null;a&&(l=a.getAttribute)||!(b=a.getAttribute("leid")));)a=a.parentNode;return b}function n(a,b,c,d,g){var e="";c||!l==b.search("<script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,7108,113,2146,307,2558,5,37,5,466,945,100,223,416,59,1,2,129,56,2,2260,4144,3956,343,3038,750,38,1524,636,3374,1769,12,2,53,12,286,1839,255,759,1085,229,4,84,640,48,1539,485,660,3,642,852,352,368,859,330,9,496,67,179,370,565,82,401,171,711,702,883,350,95,210,468,410,959,5476301,822,108,5996661,2800420,882,444,1,2,80,1,1796,1,2562,1,748,936,4,559,1,4265,1,1,2,1331,4142,2609,155,17,13,72,139,4,2,20,2,169,13,19,46,5,39,96,548,29,2,2,1,2,1,2,2,7,4,1,2,2,2,2,2,353,513,186,1,1,158,3,2,2,2,2,2,4,2,3,3,269,1601,141,82,4,490,139,1,10,53,45,34,4,38,2,2,38,2,23952036,4041690,3,450,3455,9,1435,159,1358,4726,3,955,3,1215,489,1900,1020,609,1830',kBL:'b0B6'};google.sn='webhp';google.kHL='en-IN'}};})();(function){var f=this||self;var h,k=[];function l(a){for(var b;a&&(l=a.getAttribute)||!(b=a.getAttribute("eid")));)a=a.parentNode;return b||h}function m(a){for(var b=null;a&&(l=a.getAttribute)||!(b=a.getAttribute("leid")));)a=a.parentNode;return b}function n(a,b,c,d,g){var e="";c||!l==b.search("<script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,7108,113,2146,307,2558,5,37,5,466,945,100,223,416,59,1,2,129,56,2,2260,4144,3956,343,3038,750,38,1524,636,3374,1769,12,2,53,12,286,1839,255,759,1085,229,4,84,640,48,1539,485,660,3,642,852,352,368,859,330,9,496,67,179,370,565,82,401,171,711,702,883,350,95,210,468,410,959,5476301,822,108,5996661,2800420,882,444,1,2,80,1,1796,1,2562,1,748,936,4,559,1,4265,1,1,2,1331,4142,2609,155,17,13,72,139,4,2,20,2,169,13,19,46,5,39,96,548,29,2,2,1,2,1,2,2,7,4,1,2,2,2,2,2,353,513,186,1,1,158,3,2,2,2,2,2,4,2,3,3,269,1601,141,82,4,490,139,1,10,53,45,34,4,38,2,2,38,2,23952036,4041690,3,450,3455,9,1435,159,1358,4726,3,955,3,1215,489,1900,1020,609,1830',kBL:'b0B6'};google.sn='webhp';google.kHL='en-IN'}};})();(function){var f=this||self;var h,k=[];function l(a){for(var b;a&&(l=a.getAttribute)||!(b=a.getAttribute("eid")));)a=a.parentNode;return b||h}function m(a){for(var b=null;a&&(l=a.getAttribute)||!(b=a.getAttribute("leid")));)a=a.parentNode;return b}function n(a,b,c,d,g){var e="";c||!l==b.search("<script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,7108,113,2146,307,2558,5,37,5,466,945,100,223,416,59,1,2,129,56,2,2260,4144,3956,343,3038,750,38,1524,636,3374,1769,12,2,53,12,286,1839,255,759,1085,229,4,84,640,48,1539,485,660,3,642,852,352,368,859,330,9,496,67,179,370,565,82,401,171,711,702,883,350,95,210,468,410,959,5476301,822,108,5996661,2800420,882,444,1,2,80,1,1796,1,2562,1,748,936,4,559,1,4265,1,1,2,1331,4142,2609,155,17,13,72,139,4,2,20,2,169,13,19,46,5,39,96,548,29,2,2,1,2,1,2,2,7,4,1,2,2,2,2,2,353,513,186,1,1,158,3,2,2,2,2,2,4,2,3,3,269,1601,141,82,4,490,139,1,10,53,45,34,4,38,2,2,38,2,23952036,4041690,3,450,3455,9,1435,159,1358,4726,3,955,3,1215,489,1900,1020,609,1830',kBL:'b0B6'};google.sn='webhp';google.kHL='en-IN'}};})();(function){var f=this||self;var h,k=[];function l(a){for(var b;a&&(l=a.getAttribute)||!(b=a.getAttribute("eid")));)a=a.parentNode;return b||h}function m(a){for(var b=null;a&&(l=a.getAttribute)||!(b=a.getAttribute("leid")));)a=a.parentNode;return b}function n(a,b,c,d,g){var e="";c||!l==b.search("<script nonce="NkszXHXVWc1k2n/AJX8KMA==">
    (function)
    {window.google={kEI:'GiUWys7vI5iRoATBkLwQ',kEXPI:'0,202454,3,1100079,56873,6059,206,4804,926,1390,383,246,5,1354,4013,1238,1122515,1197720,702,380069,16114,19398,9286,17572,4859,1361,9290,3026,17583,4020,978,13228,3847,4192,6430,22741,5081,885,708,1279,2742,149,1103,840,2197,4100,109,3405,606,2023,1777,520,14670,3227,2845,7,5599,11851,16320,1851,2614,13142,3,576,1014,1,5445,148,11323,2652,4,1528,656,1648,6463,576,13271,7038,1714,3050,2658,7357,30,13628,2305,2132,16786,5824,2533,4094,4052,3,3541,1,16807,38,25309,2,14022,1999,716,255,4550,744,5852,10463,1160,5679,1020,2381,2717,18298,2,2,5,7717,4568,6253,6723,9822,6879,1252,4588,2,2,5,1238,12137,2830,1539,2333,461,8,5008,1073,58,1337,445,2,2,1,6394,6104,588,754,3472,309,2,1,6159,843,153,1406,10,1,436,1047,71
```

b) Tags (title, p, a, div)

HANDWRITTEN CODE:

b) VIBHU KUMAR SINGH 19BCE0210

```
import requests
from bs4 import BeautifulSoup
URL = "https://google.com"

r = requests.get(URL).text
soup = BeautifulSoup(r, 'html.parser')

print(soup.head.title)

p_tags = soup.find_all('p')
for text in p_tags:
    print(text)

a_tags = soup.find_all('a')
for text in a_tags:
    print(text)

div_tags = soup.find_all('div')
for text in div_tags:
    print(text)
```

CODE:

```
import requests
from bs4 import BeautifulSoup
URL = "https://google.com"

r=requests.get(URL).text
soup=BeautifulSoup(r, 'html.parser')

print("\n[4mTitle Tag:[0m ")
print(soup.head.title)

print("\n[4mp Tags:[0m ")
p_tags=soup.find_all('p')
for text in p_tags:
    print(text,end="")
```

```
print('\n')

print("\n\033[4m Tags:\033[0m ")
a_tags=soup.find_all('a')
for text in a_tags:
    print(text,end="")
    print('\n')

print("\n\033[4mdiv Tags:\033[0m ")
div_tags=soup.find_all('div')
for text in div_tags:
    print(text,end="")
    print('\n')

print("\n\033[4mTitle Tag:\033[0m ")
print(soup.head.title)

print("\n\033[4mp Tags:\033[0m ")
p_tags=soup.find_all('p')
for text in p_tags:
    print(text,end="")
    print('\n')

print("\n\033[4ma Tags:\033[0m ")
a_tags=soup.find_all('a')
for text in a_tags:
    print(text,end="")
    print('\n')

print("\n\033[4mdiv Tags:\033[0m ")
div_tags=soup.find_all('div')
for text in div_tags:
    print(text,end="")
    print('\n')
```

CODE SCREENSHOT:

```
C:\Users\Vibhu\OneDrive - vit.ac.in\Desktop\Winter Semester 21-22\Web Mining\ELA\LAB-2\htmlTags.py
rawHTML.py × htmlTags.py ×
1 import requests
2 from bs4 import BeautifulSoup
3 URL = "https://google.com"
4
5 r=requests.get(URL).text
6 soup=BeautifulSoup(r, 'html.parser')
7
8 print('\n\033[4mTitle Tag:\033[0m ')
9 print(soup.head.title)
10
11 print('\n\033[4mp Tags:\033[0m ')
12 p_tags=soup.find_all('p')
13 for text in p_tags:
14     print(text,end="")
15     print('\n')
16
17
18 print('\n\033[4ma Tags:\033[0m ')
19 a_tags=soup.find_all('a')
20 for text in a_tags:
21     print(text,end="")
22     print('\n')
23
24 print('\n\033[4mdiv Tags:\033[0m ')
25 div_tags=soup.find_all('div')
26 for text in div_tags:
27     print(text,end="")
28     print('\n')
```

OUTPUT:

Title Tag:

<title>Google</title>

p Tags:

<p style="font-size:8pt;color:#70757a">© 2022 - Privacy - Terms</p>

a Tags:

Images
Maps
Play
YouTube
Sign in
Advertising Programs
Business Solutions
About Google
...

div Tags:

```
<div id="mngb"><div id="gbar"><nobr><b class="gb1">Search</b> <a class="gb1"
href="https://www.google.co.in/imghp?hl=en&tab=wi">Images</a> <a class="gb1"
href="https://maps.google.co.in/maps?hl=en&tab=wl">Maps</a> <a class="gb1"
href="https://play.google.com/?hl=en&tab=w8">Play</a> <a class="gb1"
href="https://www.youtube.com/?gl=IN&tab=w1">YouTube</a> <a class="gb1"
href="https://news.google.com/?tab=wn">News</a> <a class="gb1"
href="https://mail.google.com/mail/?tab=wm">Gmail</a> <a class="gb1"
href="https://drive.google.com/?tab=wo">Drive</a> <a class="gb1"
href="https://www.google.co.in/intl/en/about/products?tab=wh" style="text-
decoration:none"><u>More</u> »</a></nobr></div><div id="guser"
width="100%"><nobr><span class="gbi" id="gbn"></span><span class="gbf"
id="gbf"></span><span id="gbe"></span><a class="gb4"
href="http://www.google.co.in/history/optout?hl=en">Web History</a> | <a class="gb4"
href="/preferences?hl=en">Settings</a> | <a class="gb4"
href="https://accounts.google.com/ServiceLogin?hl=en&passive=true&continue=h
ttps://www.google.com/&ec=GAZAAQ" id="gb_70" target="_top">Sign
in</a></nobr></div><div class="gbh" style="left:0"></div><div class="gbh"
style="right:0"></div></div>
```

```
<div id="gbar"><nobr><b class="gb1">Search</b> <a class="gb1"
href="https://www.google.co.in/imghp?hl=en&tab=wi">Images</a> <a class="gb1"
href="https://maps.google.co.in/maps?hl=en&tab=wl">Maps</a> <a class="gb1"
href="https://play.google.com/?hl=en&tab=w8">Play</a> <a class="gb1"
href="https://www.youtube.com/?gl=IN&tab=w1">YouTube</a> <a class="gb1"
href="https://news.google.com/?tab=wn">News</a> <a class="gb1"
href="https://mail.google.com/mail/?tab=wm">Gmail</a> <a class="gb1"
href="https://drive.google.com/?tab=wo">Drive</a> <a class="gb1"
href="https://www.google.co.in/intl/en/about/products?tab=wh" style="text-
decoration:none"><u>More</u> »</a></nobr></div>
```

```
<div id="guser" width="100%"><nobr><span class="gbi" id="gbn"></span><span
class="gbf" id="gbf"></span><span id="gbe"></span><a class="gb4"
href="http://www.google.co.in/history/optout?hl=en">Web History</a> | <a class="gb4"
href="/preferences?hl=en">Settings</a> | <a class="gb4"
href="https://accounts.google.com/ServiceLogin?hl=en&passive=true&continue=h
ttps://www.google.com/&ec=GAZAAQ" id="gb_70" target="_top">Sign
in</a></nobr></div>
```

...

OUTPUT SCREENSHOT:


```
Console 1/A
Title Tag:
<title>Google</title>

p Tags:
<p style="font-size:8pt;color:#70757a">© 2022 - <a href="/intl/en/policies/privacy/">Privacy</a> - <a href="/intl/en/policies/terms/">Terms</a></p>

a Tags:
<a class="gb1" href="https://www.google.co.in/imghp?hl=en&tab=wi">Images</a>

<a class="gb1" href="https://maps.google.co.in/maps?hl=en&tab=w1">Maps</a>

<a class="gb1" href="https://play.google.com/?hl=en&tab=w8">Play</a>

<a class="gb1" href="https://www.youtube.com/?gl=IN&tab=w1">YouTube</a>

<a class="gb1" href="https://news.google.com/?tab=wn">News</a>

<a class="gb1" href="https://mail.google.com/mail/?tab=wm">Gmail</a>

<a class="gb1" href="https://drive.google.com/?tab=wo">Drive</a>

<a class="gb1" href="https://www.google.co.in/intl/en/about/products?tab=wh" style="text-decoration:none"><u>More</u> ></a>

<a class="gb4" href="http://www.google.co.in/history/optout?hl=en">Web History</a>

<a class="gb4" href="/preferences?hl=en">Settings</a>

<a class="gb4" href="https://accounts.google.com/ServiceLogin?hl=en&passive=true&continue=https://www.google.com/&ec=GAZAAQ" id="gb_70" target="_top">Sign in</a>

<a href="/advanced_search?hl=en-IN&authuser=0">Advanced search</a>

<a href="https://www.google.com/setprefs?sig=0_whoBVMoFuGNrENGQriaJv5NSiow%3D&hl=hi&source=homepage&sa=X&ved=0ahUKEWi4m4z655X2AhWJGKYKHSE8AAIQ2ZgBCAU">हिन्दी </a>

Python console History
conda: base (Python 3.7.6) Line 28, Col 16 ASCII CRLF RW Mem 82%
```

```
Console 1/A
div Tags:
<div id="mngb"><div id="gbar"><no><b class="gb1">Search</b> <a class="gb1" href="https://www.google.co.in/imghp?hl=en&tab=wi">Images</a> <a class="gb1" href="https://maps.google.co.in/maps?hl=en&tab=w1">Maps</a> <a class="gb1" href="https://play.google.com/?hl=en&tab=w8">Play</a> <a class="gb1" href="https://www.youtube.com/?gl=IN&tab=w1">YouTube</a> <a class="gb1" href="https://news.google.com/?tab=wn">News</a> <a class="gb1" href="https://mail.google.com/mail/?tab=wm">Gmail</a> <a class="gb1" href="https://drive.google.com/?tab=wo">Drive</a> <a class="gb1" href="https://www.google.co.in/intl/en/about/products?tab=wh" style="text-decoration:none"><u>More</u> ></a></div><div id="guser" width="100%"><no><span class="gbi" id="gbn"></span><span class="gbf" id="gbf"></span><span id="gbe"></span><a class="gb4" href="http://www.google.co.in/history/optout?hl=en">Web History</a> | <a class="gb4" href="/preferences?hl=en">Settings</a> | <a class="gb4" href="https://accounts.google.com/ServiceLogin?hl=en&passive=true&continue=https://www.google.com/&ec=GAZAAQ" id="gb_70" target="_top">Sign in</a></div><div class="gbh" style="left:0"></div><div class="gbh" style="right:0"></div></div>

<div id="gbar"><no><b class="gb1">Search</b> <a class="gb1" href="https://www.google.co.in/imghp?hl=en&tab=wi">Images</a> <a class="gb1" href="https://maps.google.co.in/maps?hl=en&tab=w1">Maps</a> <a class="gb1" href="https://play.google.com/?hl=en&tab=w8">Play</a> <a class="gb1" href="https://www.youtube.com/?gl=IN&tab=w1">YouTube</a> <a class="gb1" href="https://news.google.com/?tab=wn">News</a> <a class="gb1" href="https://mail.google.com/mail/?tab=wm">Gmail</a> <a class="gb1" href="https://drive.google.com/?tab=wo">Drive</a> <a class="gb1" href="https://www.google.co.in/intl/en/about/products?tab=wh" style="text-decoration:none"><u>More</u> ></a></div>

<div id="guser" width="100%"><no><span class="gbi" id="gbn"></span><span class="gbf" id="gbf"></span><span id="gbe"></span><a class="gb4" href="http://www.google.co.in/history/optout?hl=en">Web History</a> | <a class="gb4" href="/preferences?hl=en">Settings</a> | <a class="gb4" href="https://accounts.google.com/ServiceLogin?hl=en&passive=true&continue=https://www.google.com/&ec=GAZAAQ" id="gb_70" target="_top">Sign in</a></div>

<div class="gbh" style="left:0"></div>

<div class="gbh" style="right:0"></div>

<div id="lga"><br/></div>

<div class="ds" style="height:32px;margin:4px 0"><input autocomplete="off" class="lst" maxlength="2048" name="q" size="57" style="margin:0;padding:5px 8px 0 6px;vertical-align:top;color:#000" title="Google Search" value=""/></div>

<div id="gac_scont"></div>

<div style="font-size:83%;min-height:3.5em"><br/><div id="gws-output-pages-elements-homepage_additional_languages_als"><style>#gws-output-pages-elements-homepage_additional_languages_als{font-size:small;margin-bottom:24px}#SivCob{color:#3c4043;display:inline-block;line-

Python console History
conda: base (Python 3.7.6) Line 28, Col 16 ASCII CRLF RW Mem 87%
```


c) All Textual Content

HANDWRITTEN CODE:

```
c) VIBHU KUMAR SINGH 19BCE0215

import requests
from bs4 import BeautifulSoup
URL="https://google.com"

r=requests.get(URL).text
soup=BeautifulSoup(r, 'html.parser')

print('\n\033[31mText Content: \033[0m')
print(soup.get_text())
```

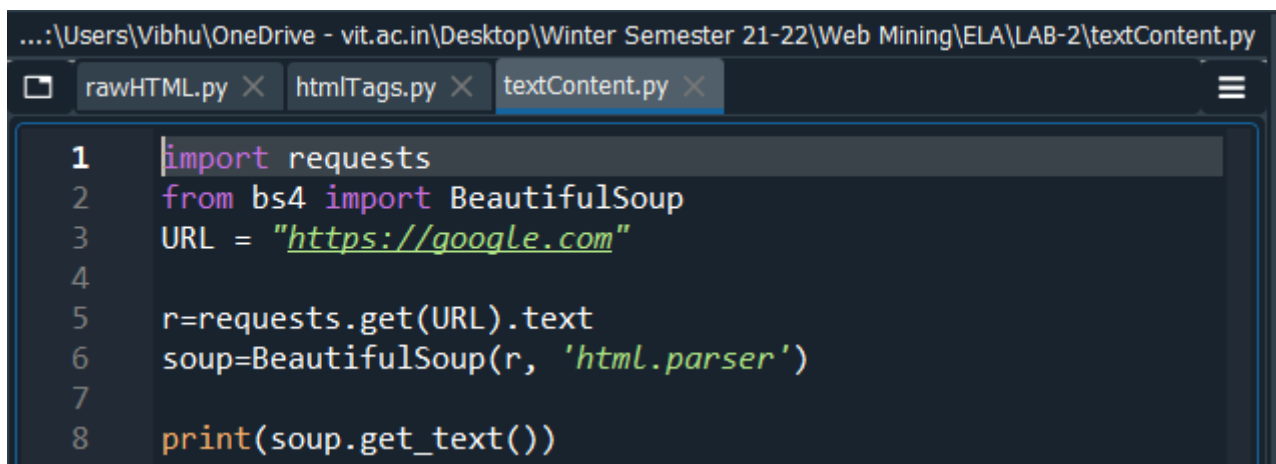
CODE:

```
import requests
from bs4 import BeautifulSoup
URL = "https://google.com"

r=requests.get(URL).text
soup=BeautifulSoup(r, 'html.parser')

print(soup.get_text())
```

CODE SCREENSHOT:



The screenshot shows a code editor window with the following code:

```
1 import requests
2 from bs4 import BeautifulSoup
3 URL = "https://google.com"
4
5 r=requests.get(URL).text
6 soup=BeautifulSoup(r, 'html.parser')
7
8 print(soup.get_text())
```

OUTPUT:

Text Content:

Google Search Images Maps Play YouTube News Gmail Drive More »Web
History | Settings | Sign in Advanced search Google offered in: हिन्दी বাংলা తెలుగు मराठी
தமிழ் ગુજરાતી ಕನ್ನಡ മലയാളം ਪੰਜਾਬੀ Advertising Programs Business Solutions About
GoogleGoogle.co.in© 2022 - Privacy – Terms

OUTPUT SCREENSHOT:

```
PS C:\Users\Vibhu\OneDrive - vit.ac.in\Desktop\Winter Semester 21-22\Web Mining\ELA\LAB-2> python -u "c:\Users\Vibhu\OneDrive - vit.ac.in\Desktop\Winter Semester 21-22\Web Mining\ELA\LAB-2\textContent.py"
```

Text Content:

GoogleSearch Images Maps Play YouTube News Gmail Drive More »Web
History | Settings | Sign in Advanced searchGoogle offered in: ह
िन्दी বাংলা తెలుగు मराठी தமிழ் ગુજરાતી ಕನ್ನಡ മലയാളം ਪੰਜਾਬੀ Adverte
ising ProgramsBusiness SolutionsAbout GoogleGoogle.co.in© 2022 -
Privacy - Terms