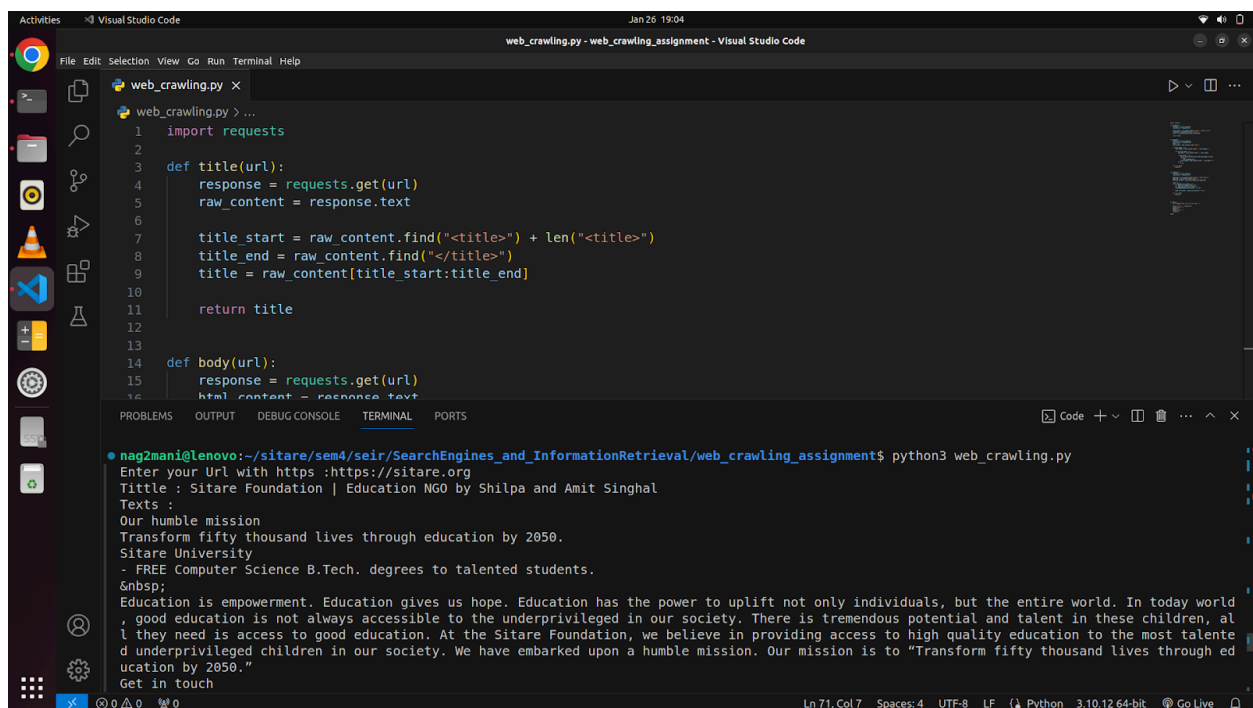


Output of my webpage Python Project using following 3 urls.

1. <https://sitare.org>

Titlle of Sitare.org;



The screenshot shows a Visual Studio Code window with a Python file named `web_crawling.py` and a terminal window displaying the output of the script.

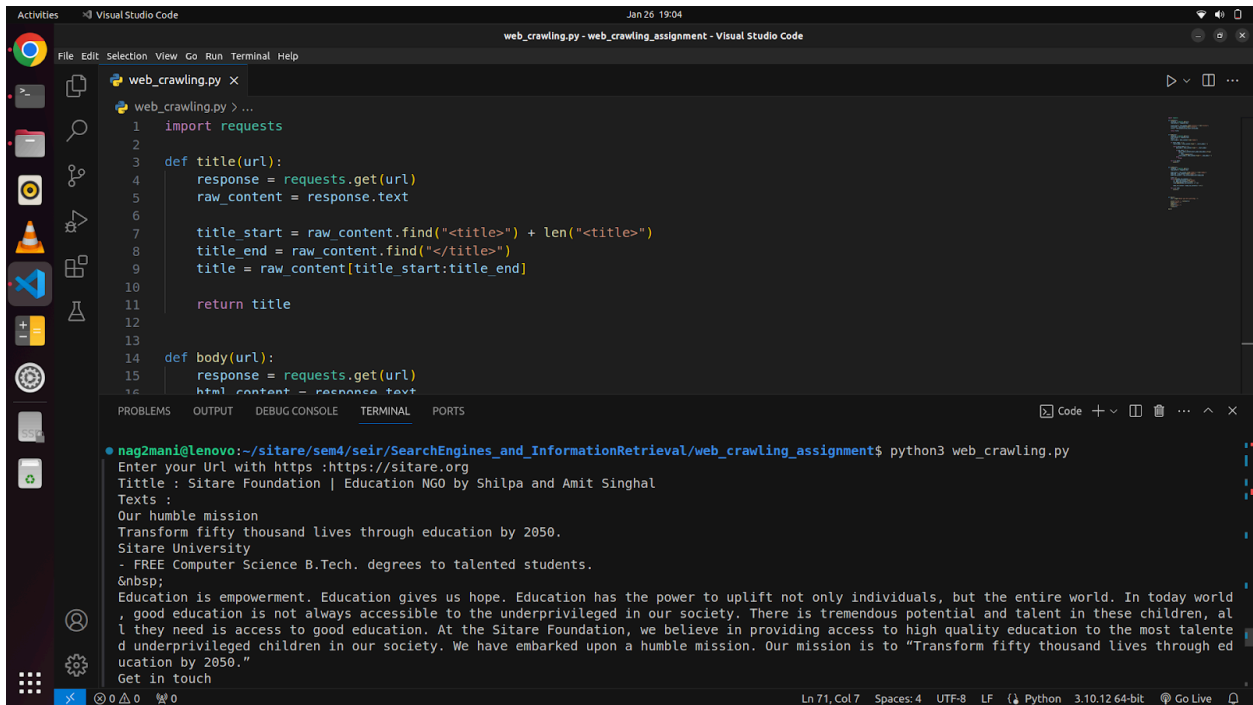
Python Code:

```
1 import requests
2
3 def title(url):
4     response = requests.get(url)
5     raw_content = response.text
6
7     title_start = raw_content.find("<title>") + len("<title>")
8     title_end = raw_content.find("</title>")
9     title = raw_content[title_start:title_end]
10
11     return title
12
13
14 def body(url):
15     response = requests.get(url)
16     html_content = response.text
```

Terminal Output:

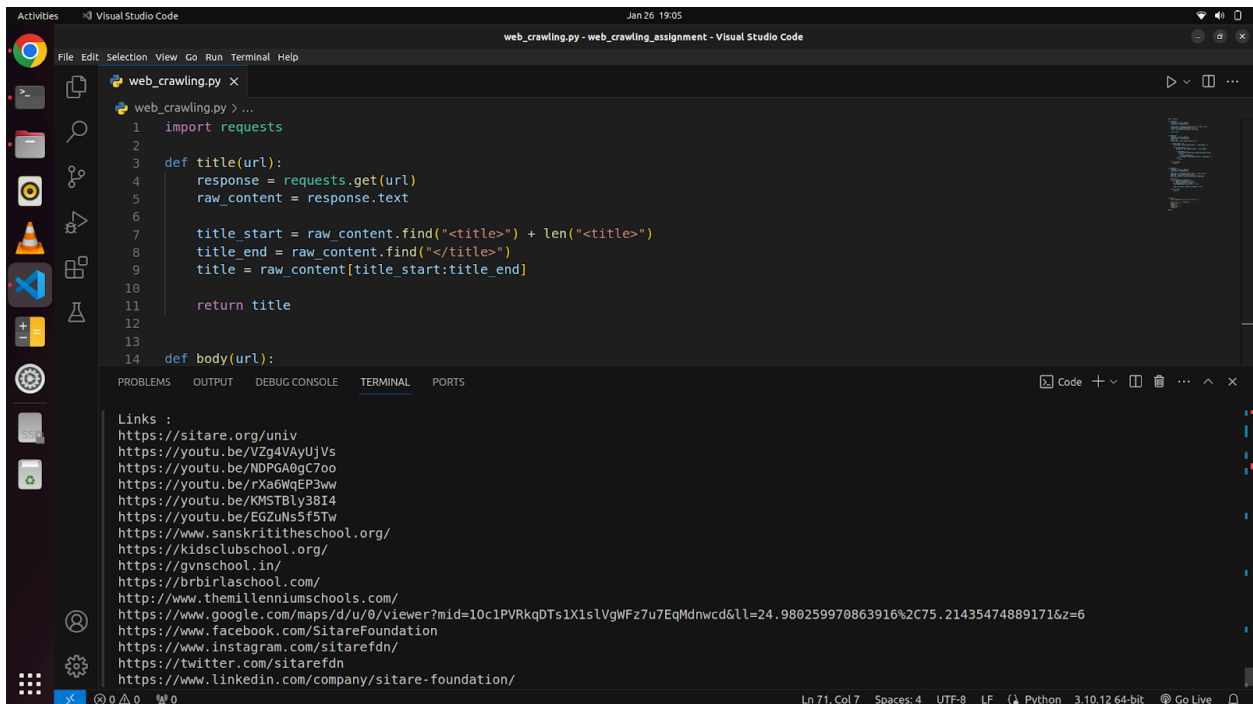
```
nag2mani@lenovo:~/sitare/sem4/seir/SearchEngines_and_InformationRetrieval/web_crawling_assignment$ python3 web_crawling.py
Enter your Url with https :https://sitare.org
Title : Sitare Foundation | Education NGO by Shilpa and Amit Singhal
Texts :
Our humble mission
Transform fifty thousand lives through education by 2050.
Sitare University
- FREE Computer Science B.Tech. degrees to talented students.
&nbsp;
Education is empowerment. Education gives us hope. Education has the power to uplift not only individuals, but the entire world. In today world
, good education is not always accessible to the underprivileged in our society. There is tremendous potential and talent in these children, al
l they need is access to good education. At the Sitare Foundation, we believe in providing access to high quality education to the most talente
d underprivileged children in our society. We have embarked upon a humble mission. Our mission is to "Transform fifty thousand lives through ed
ucation by 2050."
Get in touch
```

Texts of Sitare.org;



```
web_crawling.py - web_crawling_assignment - Visual Studio Code
1 import requests
2
3 def title(url):
4     response = requests.get(url)
5     raw_content = response.text
6
7     title_start = raw_content.find("<title>") + len("<title>")
8     title_end = raw_content.find("</title>")
9     title = raw_content[title_start:title_end]
10
11     return title
12
13
14 def body(url):
15     response = requests.get(url)
16     html_content = response.text
17
18     body_start = raw_content.find("<body>") + len("<body>")
19     body_end = raw_content.find("</body>")
20     body = raw_content[body_start:body_end]
21
22     return body
23
24 if __name__ == '__main__':
25     url = input("Enter your URL with https :https://sitare.org\n")
26     title = title(url)
27     body = body(url)
28     print("Title : Sitare Foundation | Education NGO by Shilpa and Amit Singhal\n")
29     print("Texts :")
30     print(body)
31
32 python3 web_crawling.py
Enter your URL with https :https://sitare.org
Title : Sitare Foundation | Education NGO by Shilpa and Amit Singhal
Texts :
Our humble mission
Transform fifty thousand lives through education by 2050.
Sitare University
- FREE Computer Science B.Tech. degrees to talented students.
&nbsp;
Education is empowerment. Education gives us hope. Education has the power to uplift not only individuals, but the entire world. In today world , good education is not always accessible to the underprivileged in our society. There is tremendous potential and talent in these children, al l they need is access to good education. At the Sitare Foundation, we believe in providing access to high quality education to the most talente d underprivileged children in our society. We have embarked upon a humble mission. Our mission is to "Transform fifty thousand lives through ed ucation by 2050."
Get in touch
```

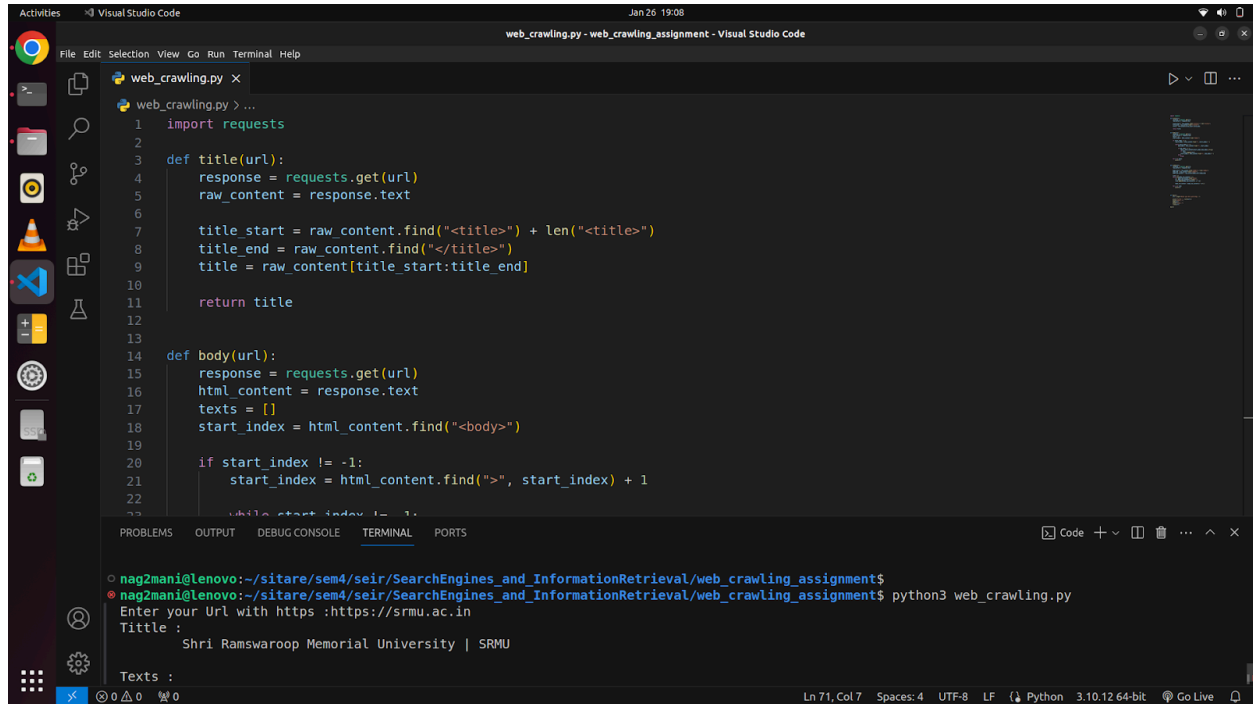
Links of sitare.org;



```
web_crawling.py - web_crawling_assignment - Visual Studio Code
1 import requests
2
3 def title(url):
4     response = requests.get(url)
5     raw_content = response.text
6
7     title_start = raw_content.find("<title>") + len("<title>")
8     title_end = raw_content.find("</title>")
9     title = raw_content[title_start:title_end]
10
11     return title
12
13
14 def body(url):
15     response = requests.get(url)
16     html_content = response.text
17
18     body_start = raw_content.find("<body>") + len("<body>")
19     body_end = raw_content.find("</body>")
20     body = raw_content[body_start:body_end]
21
22     return body
23
24 if __name__ == '__main__':
25     url = input("Enter your URL with https :https://sitare.org\n")
26     title = title(url)
27     body = body(url)
28     print("Title : Sitare Foundation | Education NGO by Shilpa and Amit Singhal\n")
29     print("Texts :")
30     print(body)
31
32 python3 web_crawling.py
Enter your URL with https :https://sitare.org
Title : Sitare Foundation | Education NGO by Shilpa and Amit Singhal
Texts :
Our humble mission
Transform fifty thousand lives through education by 2050.
Sitare University
- FREE Computer Science B.Tech. degrees to talented students.
&nbsp;
Education is empowerment. Education gives us hope. Education has the power to uplift not only individuals, but the entire world. In today world , good education is not always accessible to the underprivileged in our society. There is tremendous potential and talent in these children, al l they need is access to good education. At the Sitare Foundation, we believe in providing access to high quality education to the most talente d underprivileged children in our society. We have embarked upon a humble mission. Our mission is to "Transform fifty thousand lives through ed ucation by 2050."
Get in touch
Links :
https://sitare.org/univ
https://youtu.be/VZg4VAyUjVs
https://youtu.be/NDPGA0gC7oo
https://youtu.be/rXa6WqEP3ww
https://youtu.be/KMSTBly38I4
https://youtu.be/EGZuNs5f5Tw
https://www.sanskrititheschool.org/
https://kidsclubschool.org/
https://gvnschool.in/
https://brbirlaschool.com/
https://www.themillenniumschoools.com/
https://www.google.com/maps/d/u/0/viewer?mid=10c1PVRkqDTs1X1slVgWFz7u7EqMdnwcd&ll=24.980259970863916%2C75.21435474889171&z=6
https://www.facebook.com/SitareFoundation
https://www.instagram.com/sitarefdn/
https://twitter.com/sitarefdn
https://www.linkedin.com/company/sitare-foundation/
```

2. <https://srmu.ac.in>

Title of SRMU;

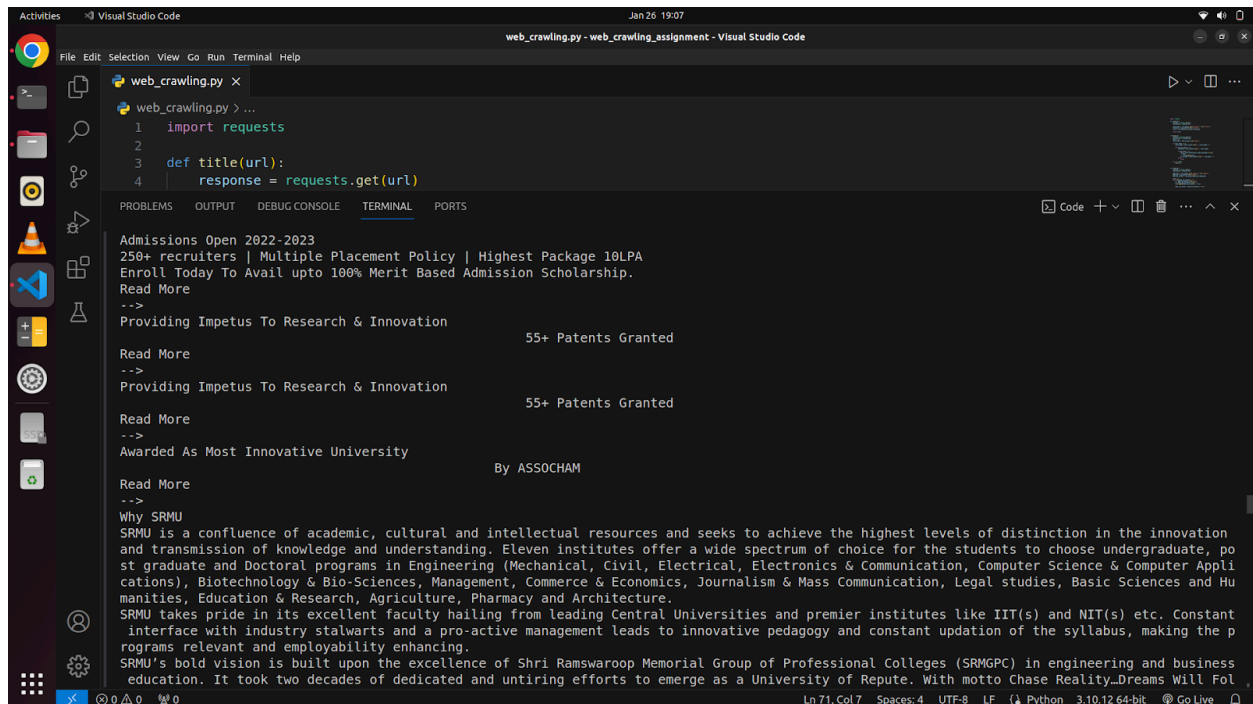


```
1 import requests
2
3 def title(url):
4     response = requests.get(url)
5     raw_content = response.text
6
7     title_start = raw_content.find("<title>") + len("<title>")
8     title_end = raw_content.find("</title>")
9     title = raw_content[title_start:title_end]
10
11     return title
12
13
14 def body(url):
15     response = requests.get(url)
16     html_content = response.text
17     texts = []
18     start_index = html_content.find("<body>")
19
20     if start_index != -1:
21         start_index = html_content.find(">", start_index) + 1
22         while start_index != -1:
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
nag2mani@lenovo:~/sitare/sem4/seir/SearchEngines_and_InformationRetrieval/web_crawling_assignment$
nag2mani@lenovo:~/sitare/sem4/seir/SearchEngines_and_InformationRetrieval/web_crawling_assignment$ python3 web_crawling.py
Enter your Url with https :https://srmu.ac.in
Title :
Shri Ramswaroop Memorial University | SRMU
Texts :
```

Texts of SRMU;

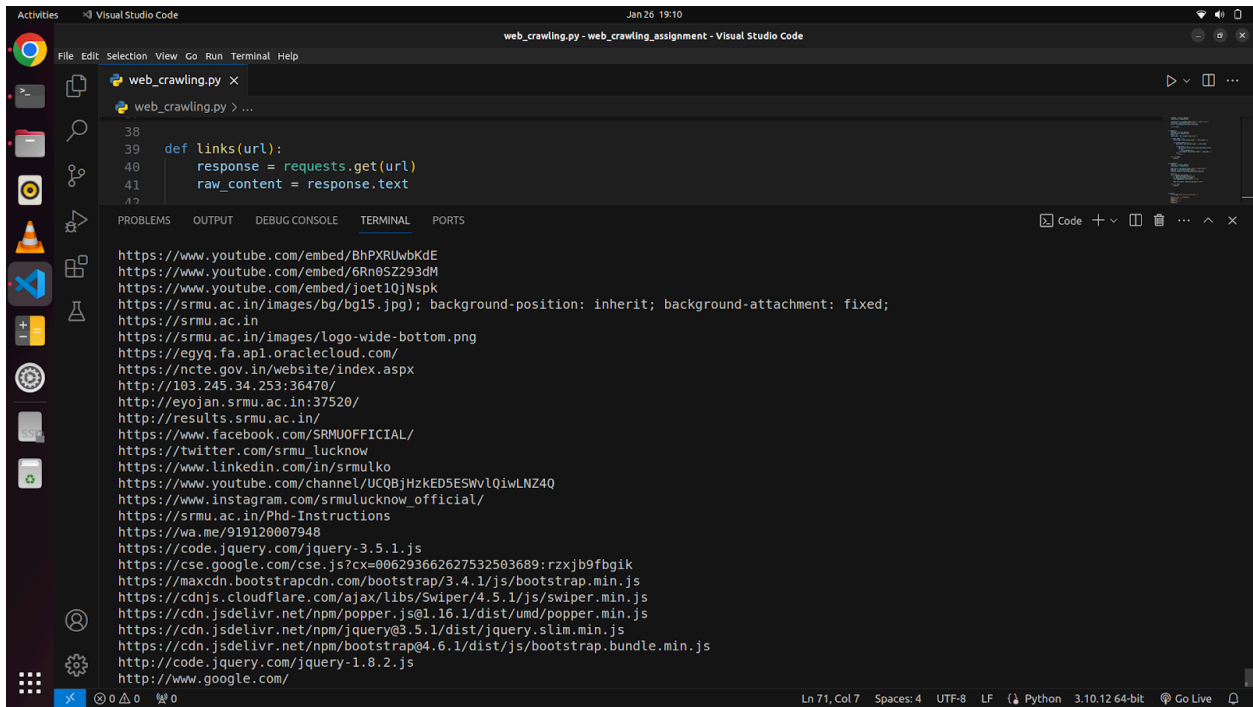


```
1 import requests
2
3 def title(url):
4     response = requests.get(url)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
Admissions Open 2022-2023
250+ recruiters | Multiple Placement Policy | Highest Package 10LPA
Enroll Today To Avail upto 100% Merit Based Admission Scholarship.
Read More
-->
Providing Impetus To Research & Innovation
55+ Patents Granted
Read More
-->
Providing Impetus To Research & Innovation
55+ Patents Granted
Read More
-->
Awarded As Most Innovative University
By ASSOCHAM
Read More
-->
Why SRMU
SRMU is a confluence of academic, cultural and intellectual resources and seeks to achieve the highest levels of distinction in the innovation and transmission of knowledge and understanding. Eleven institutes offer a wide spectrum of choice for the students to choose undergraduate, post graduate and Doctoral programs in Engineering (Mechanical, Civil, Electrical, Electronics & Communication, Computer Science & Computer Applications), Biotechnology & Bio-Sciences, Management, Commerce & Economics, Journalism & Mass Communication, Legal studies, Basic Sciences and Humanities, Education & Research, Agriculture, Pharmacy and Architecture.
SRMU takes pride in its excellent faculty hailing from leading Central Universities and premier institutes like IIT(s) and NIT(s) etc. Constant interface with industry stalwarts and a pro-active management leads to innovative pedagogy and constant updation of the syllabus, making the programs relevant and employability enhancing.
SRMU's bold vision is built upon the excellence of Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC) in engineering and business education. It took two decades of dedicated and untiring efforts to emerge as a University of Repute. With motto Chase Reality_Dreams Will Follow
```

Links of SRMU;



The screenshot shows the Visual Studio Code interface. The top editor pane displays a Python script named `web_crawling.py` with the following code:

```
38
39 def links(url):
40     response = requests.get(url)
41     raw_content = response.text
42
```

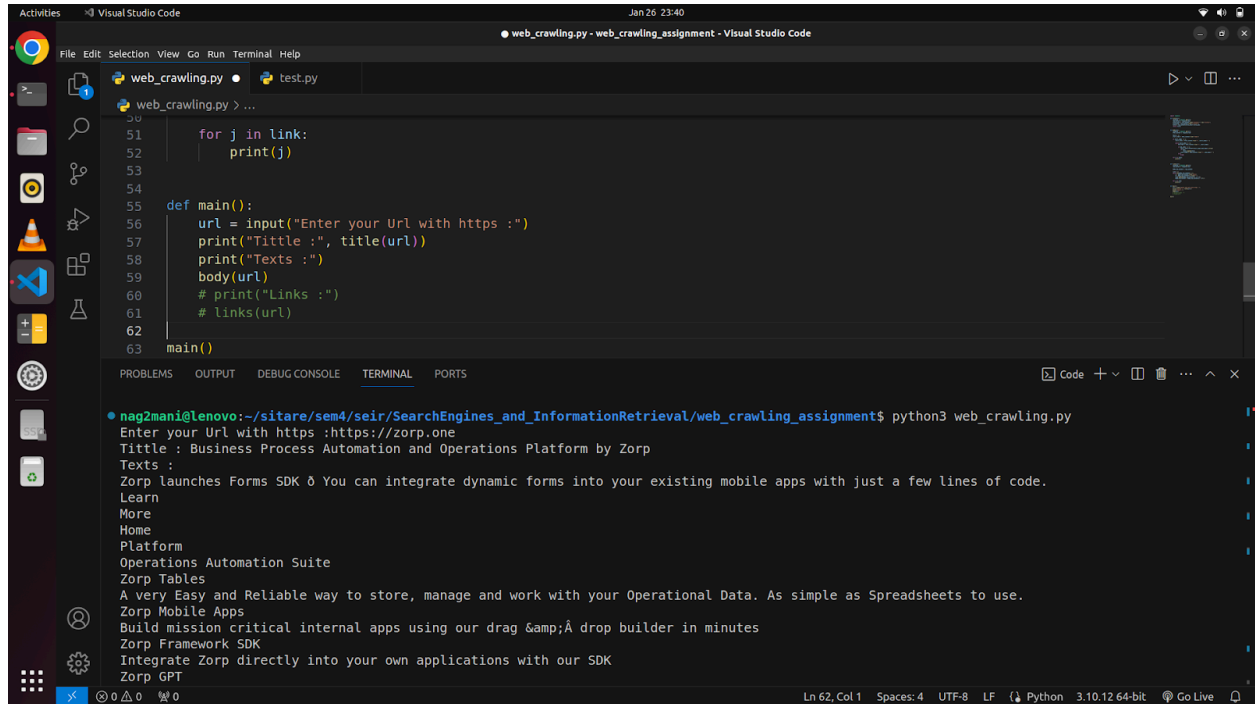
The bottom panel shows the `TERMINAL` output, which lists various URLs extracted from the website. The URLs include:

- `https://www.youtube.com/embed/BhPXRUwbKdE`
- `https://www.youtube.com/embed/6Rn0SZ293dM`
- `https://www.youtube.com/embed/joet1QjNspk`
- `https://srmu.ac.in/images/bg/bg15.jpg); background-position: inherit; background-attachment: fixed;`
- `https://srmu.ac.in`
- `https://srmu.ac.in/images/logo-wide-bottom.png`
- `https://egyq.fa.ap1.oraclecloud.com/`
- `https://ncte.gov.in/website/index.aspx`
- `http://103.245.34.253:36470/`
- `http://eyojan.srmu.ac.in:37520/`
- `http://results.srmu.ac.in/`
- `https://www.facebook.com/SRMUOFFICIAL/`
- `https://twitter.com/srmu_lucknow`
- `https://www.linkedin.com/in/srmulko`
- `https://www.youtube.com/channel/UCQBjHzKED5ESWvLQiwLNZ4Q`
- `https://www.instagram.com/srmulucknow_official/`
- `https://srmu.ac.in/Phd-Instructions`
- `https://wa.me/919120007948`
- `https://code.jquery.com/jquery-3.5.1.js`
- `https://cse.google.com/cse.js?cx=006293662627532503689:rzxjb9fbgik`
- `https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js`
- `https://cdnjs.cloudflare.com/ajax/libs/Swiper/4.5.1/js/swiper.min.js`
- `https://cdn.jsdelivr.net/npm/popper.js@1.16.1/dist/umd/popper.min.js`
- `https://cdn.jsdelivr.net/npm/jquery@3.5.1/dist/jquery.slim.min.js`
- `https://cdn.jsdelivr.net/npm/bootstrap@4.6.1/dist/js/bootstrap.bundle.min.js`
- `http://code.jquery.com/jquery-1.8.2.js`
- `http://www.google.com/`

The status bar at the bottom indicates the cursor is at line 71, column 7, with 4 spaces, UTF-8 encoding, LF line endings, Python interpreter, and 3.10.12 64-bit architecture.

3. <https://zorp.one>

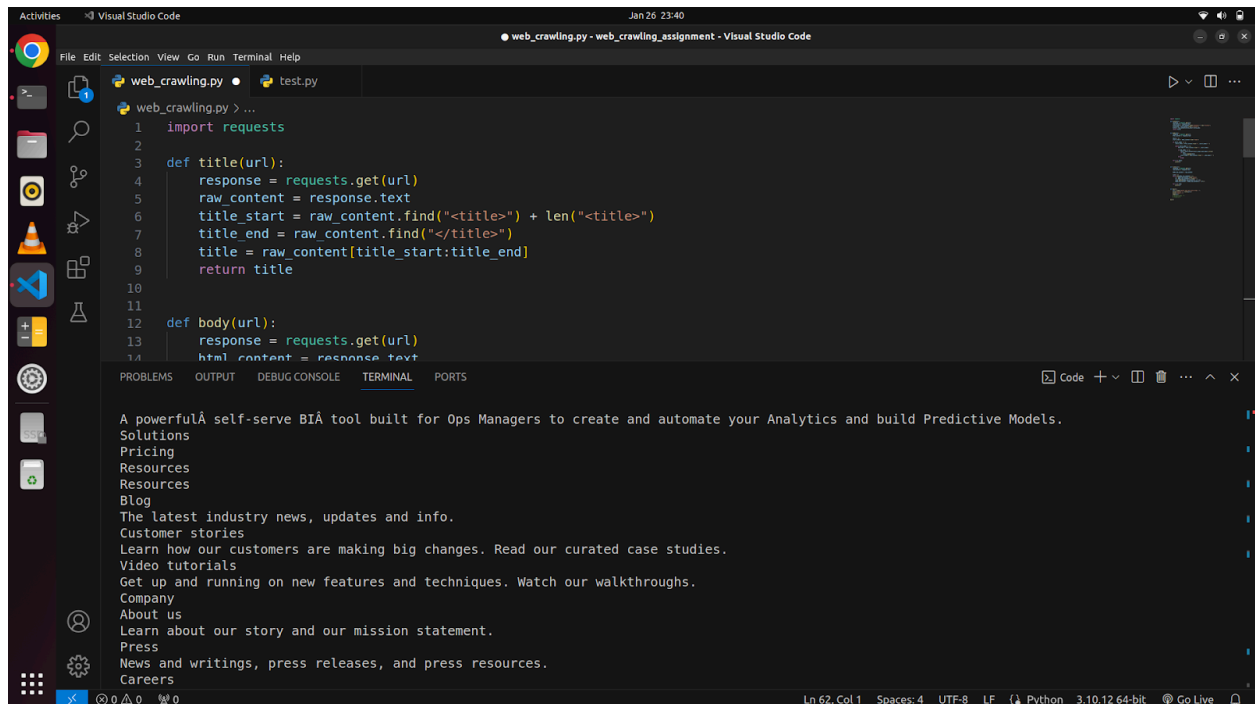
Title of ZORP;



```
web_crawling.py
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51     for j in link:
52         print(j)
53
54
55 def main():
56     url = input("Enter your Url with https :)")
57     print("Title :", title(url))
58     print("Texts :")
59     body(url)
60     # print("Links :")
61     # links(url)
62
63 main()
```

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
nag2mani@Lenovo:~/sitare/sem4/seir/SearchEngines_and_InformationRetrieval/web_crawling_assignment$ python3 web_crawling.py
Enter your Url with https :https://zorp.one
Title : Business Process Automation and Operations Platform by Zorp
Texts :
Zorp launches Forms SDK ð You can integrate dynamic forms into your existing mobile apps with just a few lines of code.
Learn
More
Home
Platform
Operations Automation Suite
Zorp Tables
A very Easy and Reliable way to store, manage and work with your Operational Data. As simple as Spreadsheets to use.
Zorp Mobile Apps
Build mission critical internal apps using our drag & drop builder in minutes
Zorp Framework SDK
Integrate Zorp directly into your own applications with our SDK
Zorp GPT
```

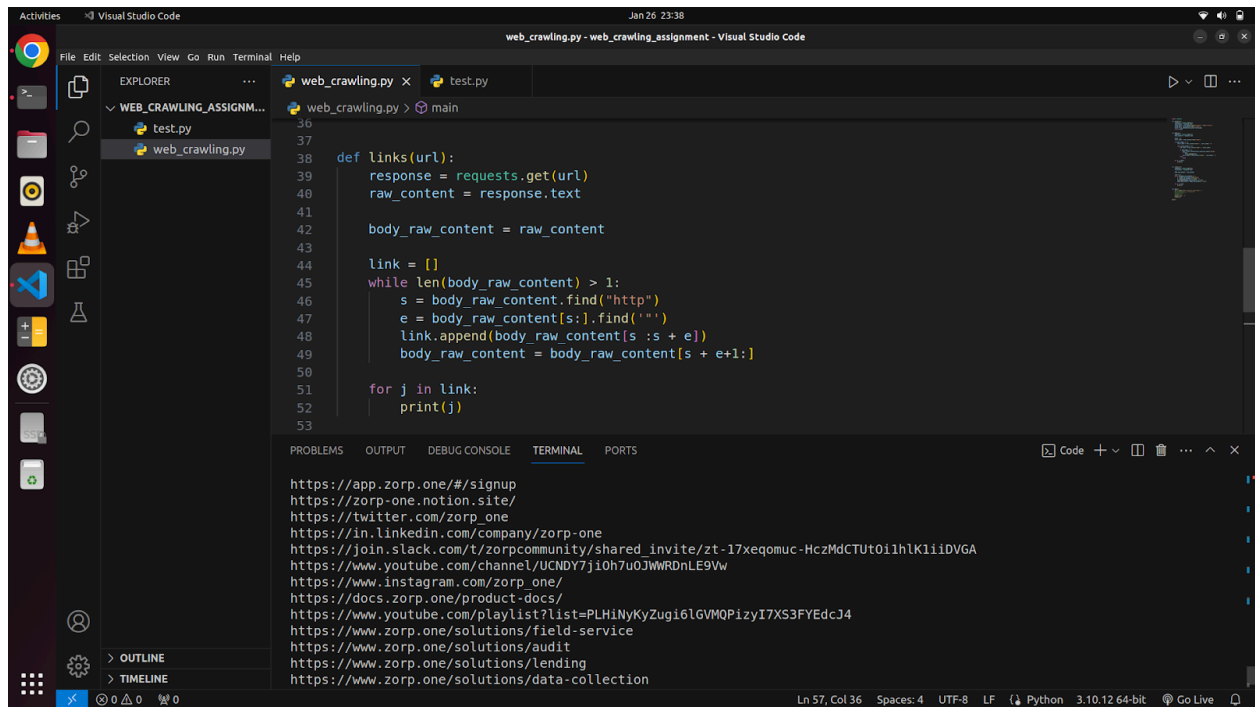
Texts of ZORP;



```
web_crawling.py
1 import requests
2
3 def title(url):
4     response = requests.get(url)
5     raw_content = response.text
6     title_start = raw_content.find("<title>") + len("<title>")
7     title_end = raw_content.find("</title>")
8     title = raw_content[title_start:title_end]
9     return title
10
11
12 def body(url):
13     response = requests.get(url)
14     html_content = response.text
```

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
A powerful self-serve BI tool built for Ops Managers to create and automate your Analytics and build Predictive Models.
Solutions
Pricing
Resources
Resources
Blog
The latest industry news, updates and info.
Customer stories
Learn how our customers are making big changes. Read our curated case studies.
Video tutorials
Get up and running on new features and techniques. Watch our walkthroughs.
Company
About us
Learn about our story and our mission statement.
Press
News and writings, press releases, and press resources.
Careers
```

Links of ZORP;



The screenshot shows the Visual Studio Code interface. The Explorer panel on the left shows a project named 'WEB_CRAWLING_ASSIGNM...' with files 'test.py' and 'web_crawling.py'. The main editor displays the 'web_crawling.py' file with the following Python code:

```
36
37
38 def links(url):
39     response = requests.get(url)
40     raw_content = response.text
41
42     body_raw_content = raw_content
43
44     link = []
45     while len(body_raw_content) > 1:
46         s = body_raw_content.find("http")
47         e = body_raw_content[s:].find('"')
48         link.append(body_raw_content[s : s + e])
49         body_raw_content = body_raw_content[s + e+1:]
50
51     for j in link:
52         print(j)
53
```

The TERMINAL panel at the bottom shows the output of the script, listing various links from ZORP One:

```
https://app.zorp.one/#/signup
https://zorp-one.notion.site/
https://twitter.com/zorp_one
https://in.linkedin.com/company/zorp-one
https://join.slack.com/t/zorpcommunity/shared_invite/zt-17xeqomuc-HczMdCTUt0i1hK1i1DVGA
https://www.youtube.com/channel/UCNDY7ji0h7u0JWWRDnLE9Vw
https://www.instagram.com/zorp_one/
https://docs.zorp.one/product-docs/
https://www.youtube.com/playlist?list=PLHiNyKyZugi6lGVMQPiZYI7XS3FYEdcJ4
https://www.zorp.one/solutions/field-service
https://www.zorp.one/solutions/audit
https://www.zorp.one/solutions/lending
https://www.zorp.one/solutions/data-collection
```

The status bar at the bottom indicates the file is at line 57, column 36, using UTF-8 encoding, LF line endings, and the Python 3.10.12 64-bit interpreter.

Github repository link for code:-

https://github.com/nag2mani/SearchEngines_and_InformationRetrieval/blob/main/web_crawling_assignment/web_crawling.py