

The image shows a Visual Studio Code editor window with a Python file named `parsing_crawler_access.py` open. The file contains a script that uses `urllib.parse` and `urllib.robotparser` to fetch and parse a robots.txt file from a website. The script defines a base URL, a user agent, and a list of paths to check. It then constructs the full URL for the robots.txt file, initializes a `RobotFileParser` object, and attempts to fetch and read the file. The output of the script is displayed in the terminal window below the editor.

```
1 from urllib.parse import urljoin
2 from urllib.robotparser import RobotFileParser
3
4 # the base URL for the site
5 site_url = "https://www.thegradcafe.com/"
6
7 # Define agent string
8 user_agent = "agentvt"
9
10 # Paths to check (
11 paths_to_check = [
12     "/", # root of the site
13     "/cgi-bin/",
14     "/admin/",
15     "/survey/", # Main survey/results path
16     "/survey/index.php",
17     "/survey/index.php?q=Computer+Science", # A specific search result page
18     "/result/986805" # An example individual result page
19 ]
20
21 # Construct the full URL for robots.txt
22 robots_txt_url = urljoin(site_url, "robots.txt")
23
24 # Initialize the RobotFileParser
25 parser = RobotFileParser()
26 parser.set_url(robots_txt_url)
27
28 try:
29     print(f"Attempting to fetch and read: {robots_txt_url}")
30     naman maadfa
```

Terminal Output:

```
PS C:\Modern Software Concepts In Python\Module 1\Module1\jhu_software_concepts\module_2> & C:\Users\Vijsay\AppData\Local\Programs\Python\Python312\python.exe "C:\Modern Software Concepts In Python\Module 1\Module1\jhu_software_concepts\module_2\parsing_crawler_access.py"
Attempting to fetch and read: https://www.thegradcafe.com/robots.txt
Successfully read robots.txt

Checking access for user agent: 'agentvt'

Can fetch '/' : True
Can fetch '/cgi-bin/' : False
Can fetch '/admin/' : True
Can fetch '/survey/' : True
Can fetch '/survey/index.php?' : True
Can fetch '/survey/index.php?q=Computer+Science?' : True
Can fetch '/result/986805?' : True

PS C:\Modern Software Concepts In Python\Module 1\Module1\jhu_software_concepts\module_2>
```