|  |  |
| --- | --- |
| **Experiment No: 7** | |
|  | |
| **Name** | Suyash Tambe |
| **PRN** | 22070126117 |
| **Date of**  **Performance** | 9nd Sept 2024 |
|  | |
| **Title** | Write a program to take URL string from the user and split the string into its 4 components |
| **Theory (short)** | URL has four components (Ref : Forouzan Book, Chapter 27) - Protocol, Host or Domain, Port and Path  For example in URL -<http://xyz.abc.com:8080/folder1/folder2/index.html>Protocol is : HTTP  Host or Domain is : xyz.abc.com  Port is : 8080  Path : /folder1/folder2/index.html  If any of the part is not there , you just display blank in that place. |
| **Program** | def extract\_url\_info(url): # Default values  protocol = domain = port = path = ""    # Split the protocol if "://" in url:  protocol, url = url.split("://", 1)  # Find the domain and port (if any) if "/" in url:  domain\_port, path = url.split("/", 1) path = "/" + path else:  domain\_port = url    # Separate domain and port if ":" in domain\_port: domain, port = domain\_port.split(":") else:  domain = domain\_port    # Output the extracted information  print(f"Protocol: {protocol}") print(f"Domain: {domain}")  print(f"Port: {port}") print(f"Path: {path}") |

|  |  |
| --- | --- |
|  | # Example usage  url = input("Enter a URL: ") extract\_url\_info(url) |
| **Output**  **Screenshots** |  |

|  |  |
| --- | --- |
|  |  |
| **Observation** | * **Protocol**: * When a URL contains a protocol (e.g., http, https, ftp), it is correctly split and identified. * If no protocol is present in the URL, the field remains empty. * **Domain**: • The domain is extracted from the URL string correctly. It includes cases where the domain is given with or without a port number. * **Port**: * If the port number is provided after the domain (e.g., :8080), it is identified and separated. * If no port is specified, the port remains blank. * **Path**: * The path is everything after the domain/port. If a path is provided (e.g.,   /path/to/page), it is correctly extracted.   * If no path is present, it remains empty. |
| **Self-**  **assessment Q&A** | NA |
| **Conclusion** | The program efficiently parses the different components of a URL (protocol, domain, port, and path) using basic string operations. If any part of the URL is missing (such as protocol, port, or path), the corresponding field remains blank, ensuring flexibility for various types of URLs. This approach allows for simple and effective URL parsing without relying on external libraries. However, for more complex cases involving query parameters, fragments, or edge cases like special characters in URLs, additional handling might be necessary. |