Assignment No. 10

Learning Outcomes:

After Completion of this experiment, students will be able to

- students will be able to understand the components of URLs.
- code to split a given URL into four components that are protocol, domain, port, and path.

Theory:

A Universal Resource Locator (URL) is a unique web address, and it represents the location of specific resources on the internet and has 4 components: The protocol (also referred to as transfer protocol or scheme) in a URL determines how data is transferred between the host and a web browser (or client), usually given as HTTP. The domain is the domain name and is always present and is the registered identification “string” (or word/phrase). The port address is not always mentioned or specified. The path refers to the exact location of a page, post, file, or other asset.

Procedure:

- Take the input of the URL.
- Check if the protocol is specified else print „NULL“ or empty.
- Then print the domain that is specified after the protocol
- Then check for the port which is after the colon and it is also either specified or unspecified.
- Finally check the path or the location.

Code:

```
url = input("Enter the URL - ")

# Split the URL by "://" to separate protocol and the rest of the URL
parts = url.split("://")

if len(parts) == 2:
    protocol, rest = parts
else:
```

```
    print("Invalid URL format")
    exit()

# Split the rest of the URL by "/" to get domain, port, and path
split_rest = rest.split("/")
domain = split_rest[0]

# Check if there is a port specified
if ":" in domain:
    domain, port = domain.split(":")
else:
    port = "Not specified"

# Get the path
path = "/".join(split_rest[1:])

print(f"Protocol - {protocol}")
print(f"Domain - {domain}")
print(f"Port - {port}")
print(f"Path - {path}")
```

Output:

```
Enter the URL - https://classroom.google.com/c/NTIzMjc3MzMyMzUz
Protocol - https
Domain - classroom.google.com
Port - Not specified
Path - c/NTIzMjc3MzMyMzUz
```

Self-Assessment

1. **What is a URL splitter program, and what is its primary purpose?**

A URL splitter program is a software tool designed to dissect a Uniform Resource Locator (URL) into its fundamental components, including the protocol (e.g., "http" or "https"), domain (the web address or IP), port (if specified), and path (the location of the resource on the server). Its primary purpose is to facilitate the manipulation, analysis, and use of URLs in various software applications, web services, and scripts.

This program is crucial for a variety of tasks:

1. URL Parsing: It dissects a complex URL into its constituent parts, allowing for easy access to individual components. This is particularly useful when applications need to interact with specific aspects of the URL.

2. Data Extraction: It provides a means to extract essential data from a URL. For instance, it can isolate the domain for establishing network connections, determine the protocol for selecting the access method, and identify the path to specify the resource's location.

3. URL Validation: It can validate URLs to ensure their correctness and adherence to expected formatting, thus minimizing errors and security risks.

4. Custom Processing: Developers can employ a URL splitter program to perform tailored operations or routing based on URL components. For example, a web server may utilize it to direct incoming requests to the relevant resource or handle specific paths differently.

In essence, a URL splitter program simplifies the management and manipulation of URLs in software, making it more straightforward to interact with web resources, build web-related applications, and automate web-related tasks.

2. Explain the basic components of a URL that a URL splitter program

A URL splitter program typically extracts the following five basic components from a URL:

1. Protocol: Specifies the communication protocol used to access the resource (e.g., "http," "https," "ftp").

2. Domain: Represents the domain name or IP address of the server hosting the resource.

3. Port: If specified, indicates the specific communication endpoint on the server; otherwise, a default port is assumed based on the protocol.

4. Path: Specifies the location of the resource on the server, including any directories or subdirectories.

5. Query Parameters: Optional key-value pairs that provide additional information to the server and modify the resource request. These parameters are included after a question mark (`?`) in the URL.

These components are essential for identifying, accessing, and interacting with web resources, and a URL splitter program breaks down the URL into these parts to make them accessible for various web-related tasks.

typically separates.

3. Describe the role of the path in a URL.

The path component in a URL serves to specify the exact location of a resource on a web server. It essentially functions as a set of directions, indicating the directory structure and subdirectories that lead to the desired resource. It plays a pivotal role in identifying the resource, helps organize content, and facilitates structured access. Additionally, it's commonly used for routing within web applications, creating bookmarks, and ensuring user-friendly, descriptive URLs that aid in search engine optimization (SEO).

4. What is the significance of the “protocol” in a URL?

The significance of the "protocol" in a URL can be summarized in four key points:

1. Communication Rules: It defines the rules and standards for data exchange between a client and a server when accessing a web resource.

2. Access Method: It specifies how the client should request and access the resource, indicating the type of service (e.g., web pages, file transfers) to be used.

3. Security Level: It influences the security of data transmission, with some protocols, like "https," providing encryption for secure communication.

4. Default Port: The protocol determines the default port number to use for the connection, ensuring compatibility and proper routing of data to the server.