# UNDERSTANDING URL

Every document on the Web has a unique address. This address is known as **U**niform **R**esource **L**ocator *URL*.

Several HTML/XHTML tags include a URL attribute value, including hyperlinks, inline images, and forms. All of them use the same syntax to specify the location of a web resource, regardless of the type or content of that resource. That's why it is known a Uniform Resource Locator.

## URL Elements

A URL is made of up several parts, each of which offers information to the web browser to help find the page. It is easier to learn the parts of a URL, if you look at the example URL given below, there are three key parts: the scheme, the host address, and the file path. The following section will discuss each of them:

```
http://www.tutorialspoint.com/index.htm
```

## The Scheme

The scheme identifies the type of protocol and URL you are linking to and therefore, how the resource should be retrieved. For example, most web browsers use Hypertext Transfer Protocol *HTTP* to pass information to communicate with the web servers and this is the reason a URL starts with http://.

There are other schemes available and you can use either of them based on your requirement:

| Scheme | Description |
|--------|-------------|
| http:// | Hypertext Transfer Protocol *HTTP* is used to request pages from Web servers and send them back from Web servers to browsers. |
| https:// | Secure Hypertext Transfer Protocol *HTTPS* encrypts the data sent between the browser and the Web server using a digital certificate. |
| ftp:// | File Transfer Protocol is another method for transferring files on the Web. While HTTP is a lot more popular for viewing Web sites because of its integration with browsers, FTP is still commonly used protocol to transfer large files across the Web and to upload source files to your Web server. |
| file:// | Used to indicate that a file is on the local hard disk or a shared directory on a LAN. |

## The Host Address

The host address is where a website can be found, either the IP address $foursetsofnumbersbetween 0 and 258$, $forexample 68.178.157.132$ or more commonly the domain name for a site such as www.tutorialspoint.com. Note that "www" is not actually part of the domain name although it is often used in the host address.

## The File Path

The filepath always begins with a forward slash character, and may consist of one or more directory or folder names. Each directory name is separated by forward slash characters and the filepath may end with a filename at the end. Here index.htm is the filename which is available in html directory:

```
http://www.tutorialspoint.com/html/index.htm
```

## Other Parts of the URL

Using credentials is a way of specifying a username and password for a password-protected part of a site. The credentials come before the host address, and they are separated from the host address by an @ sign. Note how the username is separated from the password by a colon. The following URL shows the username *admin* and the password *admin123*:

```
http://admin:admin123@tutorialspoint.com/admin/index.htm
```

Using the above URL, you can authenticate administrator and if provided ID and Password are correct then administrator will have access on index.htm file available in admin directory.

You can use a telnet URL to connect to a server as follows :

```
telnet://user:password@tutorialspoint.com:port/
```

Another important information is Web Server *Port Number*. By default HTTP Server runs on port number 80. But if you are running a server on any other port number then it can be porvided as follows, assuming server is running on port 8080:

```
http://www.tutorialspoint.com:8080/index.htm
```

*Fragment identifiers* can be used after a filename to indicate a specific part of the page that a browser should go immediately. Following is an example to reach to the top of page html_text_links.htm.

```
http://www.tutorialspoint.com/html/html_text_links.htm#top
```

You can pass some information to the server using URL. When you use a form on a webpage, such as a search form or an online order form, the browser can append the information you supply to the URL to pass information from your browser to the server as follows:

```
http://www.tutorialspoint.com/cgi-bin/search.cgi?searchTerm=HTML
```

Here, **searchTerm=HTML** is passed to the server where search.cgi script is used to parse this passed information and take further action.

## Absolute and Relative URLs

You may address a URL in one of the following two ways:

- **Absolute** - An absolute URL is the complete address of a resource. For example http://www.tutorialspoint.com/html/html_text_links.htm

- **Relative** - A relative URL indicates where the resource is in relation to the current page. Given URL is added with the <base> element to form a complete URL. For example /html/html_text_links.htm

## Reserved and Unsafe Characters

Reserved characters are those have a specific meaning within the URL. For example, the slash character separates elements of a pathname within a URL. If you need to include a slash in a URL that is not intended to be an element separator then you need to encode it as %2F:

Unsafe characters are those have no special meaning within the URL but may have a special meaning in the context the URL is written. For example, double quotes "" delimit URL attribute values in tags. If you need to include a double quotation mark directly in a URL, you would probably confuse the browser. Instead, you should encode the double quotation mark to avoid any possible conflict.

You can check **HTML URL Encoding** tutorial to understand this encoding, reserved and unsafe characters.