How does a website work

What is a website
What is a web browser
What is HTML
What is a web server

What is a website

- a collection of separate documents and files (text, graphics, PDFs, etc.) that are usually connected together in some way
- web pages are text documents which are marked up in order to be displayed in a web browser such as Google Chrome, Firefox, Microsoft Internet Explorer or Edge, or Apple's Safari

What is a web browser

- a computer program used for accessing sites or information on a network (such as the World Wide Web)
- the browser application retrieves (or fetches) code, usually written in HTML (HyperText Markup Language) and other computer languages, from a web server.
- the web address, or URL (Uniform Resource Locator) in the address bar tells the browser where to obtain a page or pages from.

What is HTML

HTML is the standard markup language for creating Web pages.

- HTML stands for Hyper Text Markup Language
- HTML describes the structure of Web pages using markup
- HTML elements are the building blocks of HTML pages
- HTML elements are represented by tags
- HTML tags label pieces of content such as "heading", "paragraph", "table", and so on
- Browsers do not display the HTML tags, but use them to render the content of the page

What is a web server

- "Web server" can refer to hardware or software, or both of them working together.
- a computer that stores web server software and a website's component files (e.g. HTML documents, images, CSS stylesheets, and JavaScript files)
- a web server includes several parts that control how web users access hosted files, at minimum an HTTP server. An HTTP server is a piece of software that understands URLs (web addresses) and HTTP (the protocol your browser uses to view webpages). It can be accessed through the domain names (like mozilla.org) of websites it stores, and delivers their content to the end-user's device.

Source code

- There are many reasons to view source code:
  - To see other developers' implementation
  - To see why the page doesn't display correctly
  - To learn from real life examples
- Google Chrome

  CTR + U

  Enter key: view-source: (i.e., view-source:https://digitalcareerinstitute.org/en/)

  or via developer tools (CTR + SHIFT + U) or click on Chrome's main menu button, located in the upper right-hand corner and represented by three vertically-aligned dots and select More Tools → Developer Tools.
- Firefox

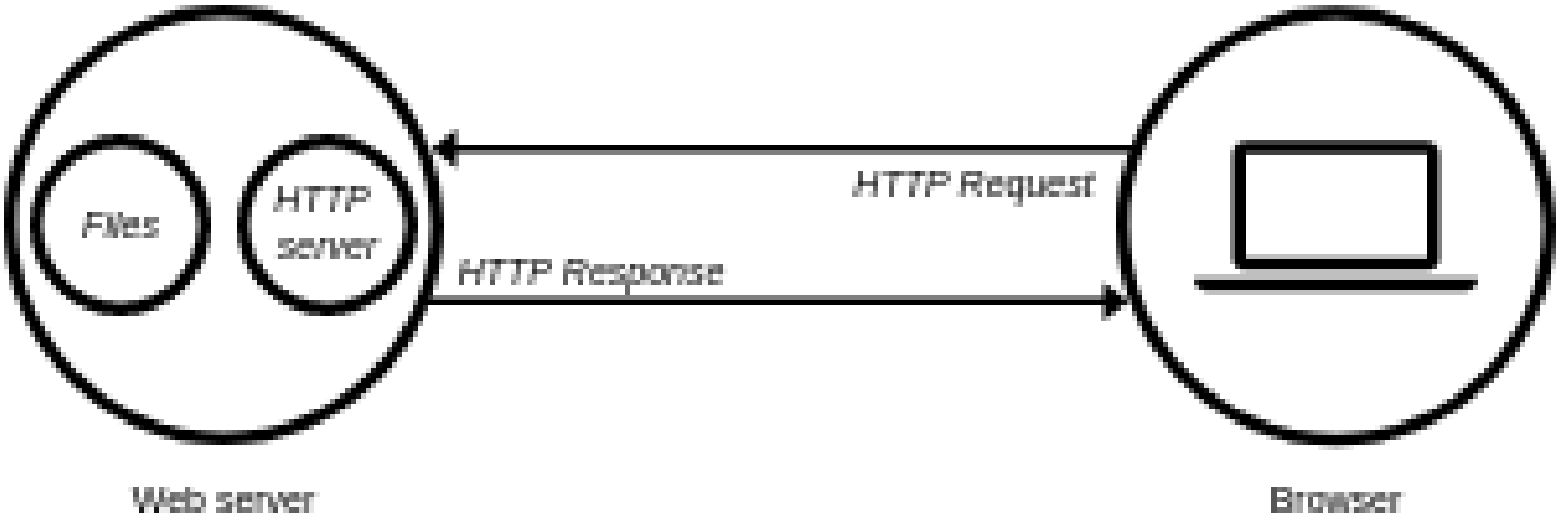
  CTRL + U (COMMAND + U on macOS)
  - view-source: (i.e., view-source:https://digitalcareerinstitute.org/en/)
  - developer tools (click on the main menu button, located in the upper right-hand corner of your browser window and represented by three horizontal lines → click developer 'wrench' icon → select 'Page source'
  - You can also view a selected area of the page by highlighting it and right clicking and selecting 'View Selection Source'

Source Code

- Microsoft Edge
  - Via Developer Tools interface. F12 or CTRL + U or click on menu button (three dots located in the upper right-hand corner) and choose the F12 Developer Tools option from the list.
- Android
  - appending the following text: *view-source:* to the front of its address (or URL) and submitting it
- IOS
  - no native methods for viewing source code using Chrome, use third party software

How does a web server works

- Whenever you type a URL in a browser requesting a file that is hosted on the web server, the browser requests the file via HTTP
- When the request reaches the correct web server (hardware), the HTTP server (software) accepts request, finds the requested document (if it doesn't then a 404 response is returned), and sends it back to the browser, also through HTTP.



Static and dynamic web server

- A static web server, or stack, consists of a computer (hardware) with an HTTP server (software). It is called "static" because the server sends hosted files "as-is" to your browser.

- A dynamic web server consists of a static web server plus extra software, most commonly an application server and a database. It is called "dynamic" because the application server updates the hosted files before sending them to your browser via the HTTP server.

Server software

- The first web servers only supported static files, such as HTML (and images), but now they most commonly allow embedding of server side applications.
- https://en.wikipedia.org/wiki/Comparison_of_web_server_software
- W3Techs reports a breakdown of market share (as of May 2014):

  Apache: 60.6%

  NGINX: 20.6%

  IIS: 13.9%

  LiteSpeed: 2.0%

  Google Servers: 1.3%
- Mongoose is developed by Cesanta Software.

Hosting files

- Usually the web pages are hosted on a dedicated web server (though it is possible to host it on your own computer)
  - is always up and running
  - is always connected to the Internet
  - has the same IP address all the time (not all ISPs provide a fixed IP address for home lines)
  - is maintained by a third-party provider

HTTP (Hypertext Transfer Protocol)

- The HTTP protocol is a request/response protocol. It specifies how to transfer hypertext between two computers

- A Protocol is a set of rules for communication between two computers. HTTP is a
  1. textual
     All commands are plain-text and human-readable.
  2. stateless protocol
     Neither the server nor the client remember previous communications.

HTTP

- https://www.w3.org/Protocols/rfc2616/rfc2616.html
- Example session

  a sample conversation between an HTTP client and an HTTP server running on www.example.com, port 80, all the data is sent in a plain-text (ASCII) encoding

  Client request:

```
GET /index.html HTTP/1.1
Host: www.example.com
```

HTTP - Server response:

```
HTTP Status for: "https://digitalcareerinstitute.org/en/"
HTTP/1.1 200 OK
Date: Wed, 28 Mar 2018 10:46:40 GMT
Server: Apache
Last-Modified: Wed, 28 Mar 2018 10:27:02 GMT
ETag: "a9c4-5687672a9eafa"
Accept-Ranges: bytes
Content-Length: 43460
Vary: Accept-Encoding,User-Agent
Connection: close
Content-Type: text/html; charset=UTF-8

<!DOCTYPE html>
<html lang="en-GB" prefix="og: http://ogp.me/ns#">
<head>
   <title>Home - Digital Career Institute</title>
</head>
<body>
  ...
</body>
</html>
```

Determines if a cached page exists

the web server will close the TCP connection immediately after the transfer of this response

HTTP

A client request (consisting in this case of the request line and only one header field) is followed by a blank line, so that the request ends with a double newline, each in the form of a carriage return followed by a line feed. The "Host" field distinguishes between various DNS names sharing a single IP address, allowing name-based virtual hosting. While optional in HTTP/1.0, it is mandatory in HTTP/1.1

HTTP Response

Stateless - minimizes the amount of data that needs to be transferred

to track the user's progress from page to page, for example when a web server is required to customize the content of a web page for a user. Solutions for these cases include:

- the use of HTTP cookies.
- server side sessions,
- hidden variables (when the current page contains a form), and
- URL-rewriting using URI-encoded parameters, e.g., /index.php?session_id=some_unique_session_code.

HTTP and HTTPS

Hyper Text Transfer Protocol Secure (HTTPS) is the secure version of HTTP (port 80), the protocol over which data is sent between your browser and the website that you are connected to. The 'S' at the end of HTTPS (port 443) stands for 'Secure'. It means all communications between your browser and the website are encrypted. This helps keep communications secure. Some search engines (e.g. Google) now give preference to sites with HTTPS.

URLs

- URL is an abbreviation that stands for "Universal Resource Locator."
- easy for people to remember and to use
- web browser finds web pages using an IP, or Internet Protocol. The IP is a series of numbers.
- DNS (Domain Name Server) translates the URL to the corresponding IP
- 193.46.215.134
- nslookup, https://www.whatismyip.com/
- Public IP (file server or website) , private IP

URL structure

- The protocol identifier and
  - It (usually, but not always) starts with "http://" or "https://" other protocols: ftp, gopher,file
- The resource name
  - host name (the name of the machine the resource lives on), filename (the pathname to the file on the machine), port number (the port number to connect to, optional), reference (a reference to a named anchor within a resource; usually identifies a specific location within a file,optional)
- Typical URL
  - It is often followed by "www"
  - Followed by the name of the website you want to visit
  - then by specific directories where the information you want to read is stored, separated by /
  - and finally, the location of the page you want to read.
  - Parts of a URL are case sensitive

Front end - Client-side Environment

The client-side environment used to run scripts is usually a browser. The processing takes place on the end users computer. The source code is transferred from the web server to the users computer over the internet and run directly in the browser.

The scripting language needs to be enabled on the client computer. Sometimes if a user is conscious of security risks they may switch the scripting facility off. When this is the case a message usually pops up to alert the user when script is attempting to run.

Scripts that execute in client side. In context of websites, it is scripts that execute in the browser of the user.

Eg: Javascript, VB etc.

(JQuery, DOJO are libraries built on top of Javascript so also client side.)

Back end - Server-side Environment

The server-side environment that runs a scripting language is a web server. A user's request is fulfilled by running a script directly on the web server to generate dynamic HTML pages. This HTML is then sent to the client browser. It is usually used to provide interactive web sites that interface to databases or other data stores on the server.

This is different from client-side scripting where scripts are run by the viewing web browser, usually in JavaScript. The primary advantage to server-side scripting is the ability to highly customize the response based on the user's requirements, access rights, or queries into data stores.

Server Side:

Scripts that execute in the Server. In context of website, it is scripts that execute on application servers.

Eg: PHP, Python, Ruby etc

Static vs dynamic content

- a server can serve either static or dynamic content. "Static" means "served as-is". Static websites are the easiest to set up
- "Dynamic" means that the server processes the content or even generates it on the fly from a database. This solution provides more flexibility, but the technical stack becomes more difficult to handle, making it dramatically more complex to build the website.

What do you need to create a website

- Text editor
- FTP or other method of uploading files to your server
- Browser

  There are dozens of browser options for your personal use, but when you're developing a website you should test it at least with the following major browsers, to make sure your site works for most people:

  - Mozilla Firefox
  - Google Chrome
  - Microsoft Internet Explorer
  - Apple Safari

  For testing consider Browsershots or Browserstack.

Sources

- https://developer.mozilla.org/en-US/docs/Learn/HTML/Introduction_to_HTML/HTML_text_fundamentals
- https://developer.mozilla.org/en-US/docs/Learn/Common_questions/
- https://developer.mozilla.org/en-US/docs/Learn/HTML/Introduction_to_HTML/Getting_started