Université Paris Cité

Lecture 2

WEB SERVER

Questions + test = 20 min

Dr. Alla Jammine
alla.jammine@u-paris.fr

What Does it Mean to Host a Website?

Definitions:

- **Server**
- **Web server**
- **Requests**
- **Respond**
- **Examples**

What are web servers used for?

Web servers are primarily used to process and manage HTTP/HTTPS requests and responses from the client system as well as:

- **Store and protect website data**
- **Control bandwidth to regulate network traffic**
- **Server-side web scripting**
- **Virtual hosting**

How web servers work

The end user processes a request via a web browser installed on a web server.

The communication between a web server or browser and the end user takes place using Hypertext Transfer Protocol (HTTP).

The primary role of a web server is to store, process, and deliver requested information or webpages to end users.

- **Physical Storage**
- **Web browser**

Web server

The term *web server* can refer to hardware or software, or both of them working together.

On the hardware side, a web server is a computer that stores web server software and a website's component files

- HTML documents, images
- CSS stylesheets
- JavaScript files.

The server delivers these files to the client’s device.

It is connected to the internet and can be accessed through a URL such as https://www.google.com/.

A web server connects to the Internet and supports physical data interchange with other devices connected to the web.

On the software side, a web server includes several parts that control how web users access hosted files. At a minimum, this is an *HTTP server*.

An HTTP server is software that understands URLs (web addresses) and HTTP (**Hypertext Transfer Protocol -** the protocol your browser uses to view webpages).

An HTTP server can be accessed through the **domain names** of the websites it stores, and it delivers the content of these hosted websites to the end user's device.

A server with just the HTTP server component is referred to as a *static* server, as opposed to a *dynamic* server which has several additional components.

A domain name - is a string of text that maps to an alphanumeric IP address, used to access a website from client software. In plain English, a domain name is the text that a user types into a browser window to reach a particular website. For instance, the domain name for Google is 'google.com'.

Who manages domain names?

Domain names are all managed by domain registries, which delegate the reservation of domain names to registrars (organizations that manage top-level domains (TLDs) '.com' and '.net' — specifically by maintaining the records of which individual domains belong to which people and organizations).

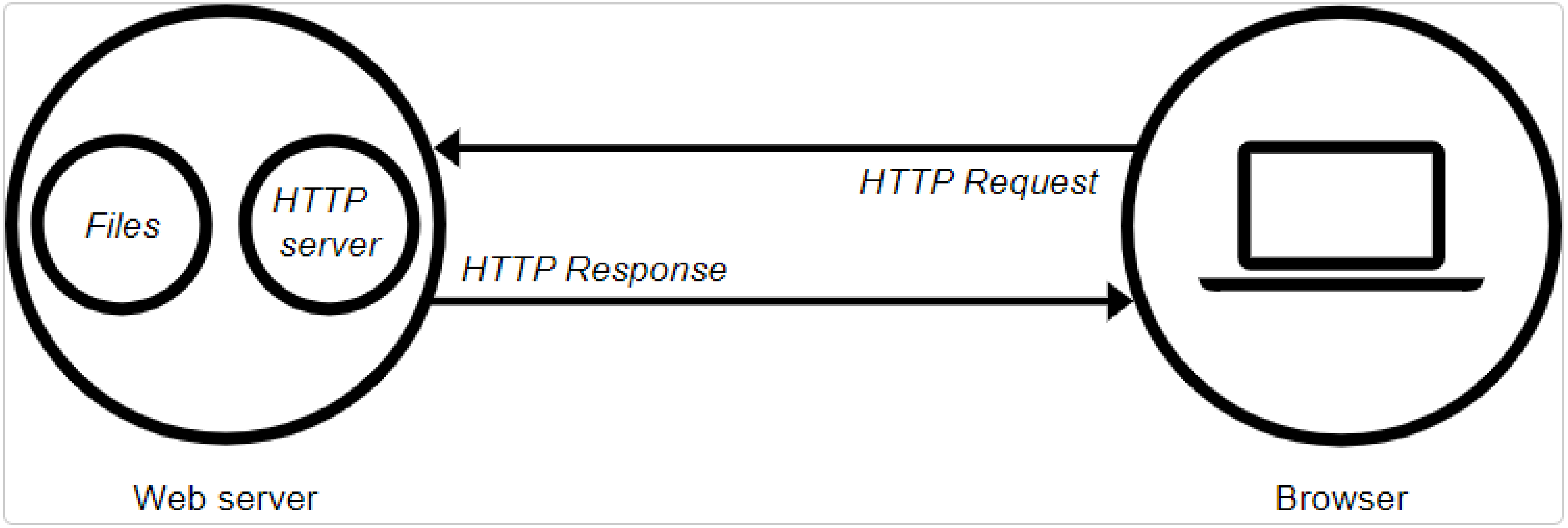
Anyone who wants to create a website can register a domain name with a registrar, and there are currently over 300 million registered domain names.

What is the difference between a domain name and a URL?

A uniform resource locator (URL), sometimes called a web address, contains the domain name of a site as well as other information, including the protocol and the path.

For example, in the URL 'https://cloudflare.com/learning/', 'cloudflare.com' is the domain name, while 'https' is the protocol and '/learning/' is the path to a specific page on the website.

Client-server communication through HTTP



Static vs. dynamic servers

To publish a website, you need either a static or a dynamic web server.

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

A **dynamic web server** consists of a static web server + extra software, most commonly an *application server* and a *database*.

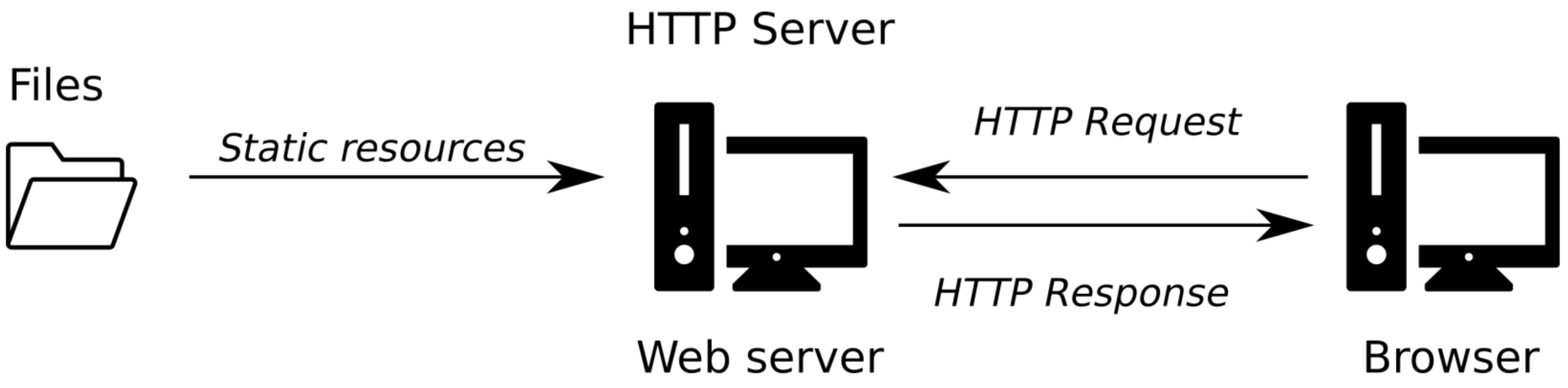
It is "dynamic" because the application server updates the hosted files before sending content to your browser via the HTTP server.

Static servers

A **static server** consists of a computer (hardware) with just an HTTP server (software).

The static server can only respond to GET requests for pre-existing HTML documents, or other types of files, and send those documents to the browser.

While loading the HTML document, the browser may send further GET requests for other pre-existing files linked in the HTML code, such as CSS, JavaScript, images, and so on.

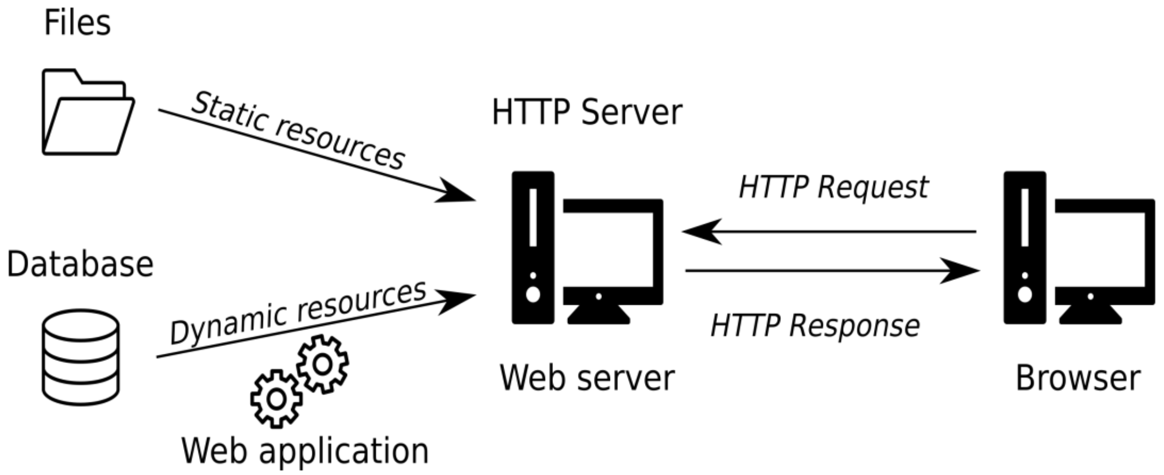


Dynamic servers

A **dynamic server** consists of an HTTP server plus extra software, most commonly an application server and a database.

Typically, the dynamic server uses its application server, i.e., software running server-side scripts, HTML templates, and a database, to assemble the HTML code.

Once assembled, the dynamically assembled HTML content is sent via HTTP, just like static content.



Communicating through HTTP
Web protocols and HTTP.

HTTP is a protocol specifying the way that communication between the client and the server takes place.

As its name—Hypertext Transfer Protocol—implies, HTTP is mainly used to transfer hypertext (i.e., HTML documents) between two computers.

HTTP is not the only protocol in use for communication on the web. For example, **FTP** and **Web socket** are examples of other web communication protocols.

However, HTTP is the most basic and most commonly used protocol.

Almost everything we do online is accomplished through HTTP communication.

The secured version of HTTP, known as **HTTPS**, is becoming a very common alternative to HTTP and thus should be mentioned.

However, HTTPS just adds a layer of security through encrypted communication and is not fundamentally different from HTTP.

In the context of the web, a protocol is a set of rules for communication between two computers.

HTTP, specifically, is a textual and stateless protocol:

- **Textual** means that all commands are plain-text, therefore human-readable.
- **Stateless** means that neither the server nor the client remember previous communications.

For example, an HTTP server, relying on HTTP alone, cannot remember if you are "logged-in" with a password, or in what step you are in a purchase transaction.

Furthermore, **only clients can make HTTP requests**, and then only to servers. **Servers can only respond to a client's HTTP request**.

HTTP methods.

The HTTP protocol defines several *methods*, or verbs, to indicate the desired action on the requested resource on the server.

The two most commonly used HTTP methods, for a request-response between client and server:

- GET
- POST

GET—Used to **request** data from the server
POST—Used to **submit** data to be processed on the server

There are a few other methods, such as PUT and DELETE, which are used much more rarely.

The GET method

The GET method is used to request data.

For example, typing a URL in the address bar of the browser in fact instructs the browser to send a GET request to the respective server.
A static server is sufficient for processing GET requests in case the requested file is physically present on the server.

The response is usually an HTML document, which is then displayed in the browser.

In addition to manual typing in the browser address bar, GET requests can also be sent programmatically, by running code.

The POST method

The POST method is used when the client sends data to be processed on the server.

It is more rarely used compared to GET, and somewhat more complicated.

For example, there is no way to send a POST request by simply typing a URL in the browser address tab, unlike with GET. Instead, making a POST request to a web server can only be made through code, such as JavaScript code.

Also, a dynamic server is required to process POST requests, where server-side scripts determine what to do with the received data.

POST requests are preferred over GET requests when we need to send substantial amounts of data to the server.

Software

An HTTP server is included in many software packages and libraries, and it does not require any special installation or configuration.

There are also several **free cloud-based** options to have a **managed static server** such as **GitHub Pages**

There are also professional HTTP server software packages, used for building both static and dynamic servers. Two most commonly used ones are **Apache HTTP Server** and **Nginx**.

Web server software list

Some of the most common web :

- **Linux web server software**
- **NGINX web server software**
- **Apache web server software**
- **IIS web server software**

Web server vs. application server differences

- **Web server:** The web server accepts and processes requests from end users for static website content. It handles requests and responses via HTTP only. Web servers are generally helpful in serving static content or static HTML webpages. It consumes fewer resources such as CPU or memory compared to the application server and provides a runtime environment for web applications.

- **Application server:** The application server can deliver web content and dynamic content required for displaying decision support, transaction results, or real-time analytics. However, its primary role is to enable interaction between the end user and server-side application code. These servers enhance interactive content or website components depending on the request. Application servers use web containers. These servers use more resources compared to web servers and provide the runtime environment for enterprise applications. These servers also support HTTP and RPC/RMI protocols.

Practical considerations

There are advantages and disadvantages to both the static and the dynamic server approaches.

Static sites (i.e., sites served with a static server) are simple, fast, and cheap, but they are harder to maintain (if they are complex) and impersonal.

Dynamic sites provide more flexibility and are easier to modify, but also slower, more expensive, and technically more difficult to build and handle.

URLs and file structure

URLs and index.html

A static server is associated with a directory on the computer, serving its contents over the web.

Once the server is running, a client can request any HTML document (or other type of file) which is located inside that directory, or any of its sub-directories, by entering a **Uniform Resource Locator (URL)** in the browser address bar.

To construct the right URL for accessing a given HTML document (or other resource), one needs to know two things:

The **IP address** of the host computer and
the **port** where the server is running,

or, alternatively, the **domain name**.

Let’s go over the separate components this URL is composed of:

http://
means we are communicating using **HTTP**. This part is automatically completed by the browser, so it can be omitted.

159.89.13.241

is the **IP address** of the web server that hosts the website. The IP address is a unique identifier given to a computer connected to a network, making it possible to identify the computer in the network.

:8000

is the **port** number where the server is running.
When using the default port for a given communication protocol, which is 80 for HTTP and 443 for HTTPS, the port number can be omitted.

/web-mapping/web-servers.html
is the location of the **document**.

With a static server, this means that within the directory we are serving there is a sub-directory named web-mapping, and inside it there is an HTML document named web-servers.html.

If you delete the last name of the file you still should see the index.html page even though it wasn't not specify any HTML file name.

This happens because standard protocol dictates that a file named index.html will be provided by default when we navigate to a directory, rather than an HTML document, on the web server.

The index.html file usually contains the first page users see when navigating to a website.

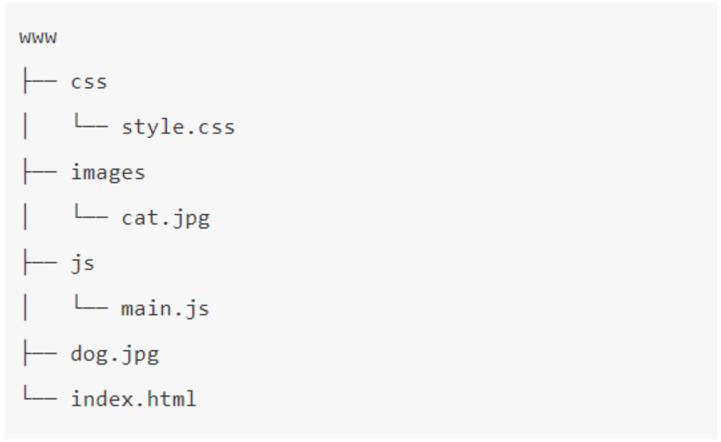
Usually navigation goes to a *textual* URL such as https://www.google.com, rather than a *numeric* IP address and port number, such as http://216.58.198.68:80/.

A **Domain Name Server (DNS)** - uses its resources to resolve the domain name into the IP address for the appropriate web server.

This saves us the trouble of remembering IP addresses, using more recognizable textual addresses instead.

File structure

The following diagram shows a hypothetical file structure of a static website directory:



```
www
├── css
|   └── style.css
├── images
|   └── cat.jpg
├── js
|   └── main.js
├── dog.jpg
└── index.html
```

The various files that comprise the website are located either in the **root directory** or in **sub-directories** within the root directory.

In the above example, www represents the root directory of the website.

In this example, the root directory www contains a default index.html document.

This means that when we browse to the directory address without specifying a file name, the index.html file is sent by default.

The root directory www also contains sub-directories.

Here, the sub-directories are used for storing additional files linked to the hypothetical HTML code of index.html:

- css—for CSS files (.css)
- images—for images
- js—for JavaScript files (.js)

It is usually convenient to have the type of sub-directory structure as shown above, where files of different types are stored in separate sub-directories.

That way, the various components that make up our website can be easier to track and maintain.

Relative paths

In the above file structure example, the images folder contains an image file named cat.jpg.

In case we want this image to be displayed on the index.html page, the HTML code in index.html needs to include an <img> element. The src attribute of that <img> element needs to refer to the cat.jpg file. Either one of the following versions will work:

<img src="/images/cat.jpg"**>** **<img src**="images/cat.jpg"**>**

Both versions of the src attribute value are known as *relative* file paths, because they are relative to a given location on the server:

In the first case, the path is relative to the **root** directory of the server, specified by the initial / symbol.

In the second case, the path is relative to the **current** directory where the HTML is loaded from, so the path just starts with a directory or file name in the same location (without the / symbol).

CSS and JavaScript

CSS and JavaScript code can be loaded from separate files, usually ending in .css and .js, respectively.

This approach is preferred to using embedded CSS or JavaScript code.

Keeping CSS and JavaScript code in separate files takes a little more effort than embedding it directly in the HTML document, but saves work as our sites become more complex, because:

1. Instead of repeating the same CSS and JavaScript code in different pages of the website, we can load the same file in all pages.

2. When modifying our external CSS or JavaScript code, all web pages loading those files are immediately affected.

Linking CSS

Linking an external CSS file can be done using the <link> element within the <head> of the HTML document.

For example, in our hypothetical static server file structure the **css** folder contains a **style.css file.**

This CSS file can be linked to the index.html document be including the following <link> element:

<link rel="stylesheet" **href**="/css/style.css"**>**

In this case, it makes sense to use a ***relative path*** (a relative file path points to a file relative to the current page) which is relative to the root directory, since a website usually has a single set of CSS and JavaScript files.

That way, exactly the same <link> element can be embedded in all HTML documents, regardless of where those HTML documents are placed.

Linking JavaScript

Linking a JavaScript code file can be done by adding a file path in the **src** attribute of a **<script>** element.

The <script> element can then be placed in the <head> or the <body> of the HTML document.

For example, our hypothetical file structure has a folder named js with a JavaScript file named main.js.

That script can be loaded in the ***index.html*** document by including the following <script> element:

<script src="/js/main.js"**></script>**

Again, a relative path, relative to the root directory, is being used.

Running a static server

So far we have discussed several background topics related to running a static server:

- Communication through HTTP
- Difference between static and dynamic servers
- Components of a URL and the file structure on the server

What is left to be done is actually *running* a server, to see how it all works in practice.

We will experiment with running a static web server using two different methods:

- A **local server**, using your own computer and Python
- A **remote server**, using the GitHub Pages platform

Remote with GitHub Pages

Python’s HTTP server is simple enough to start working with, but there are other difficulties if you intend to use it for **production**, i.e., in real-life scenarios, where stability is essential.

For instance, you need to take care of the network administration issues such as making sure your server has an IP address that can be reached from other computers, i.e., a public IP address, and that the server is not behind a firewall.

In addition, you need to make sure the IP address of the computer always stays the same (a static IP address), take care of the hardware of your server, such as making sure the computer is always running and connected to the internet, make sure that the server is restarted in case the computer restarts, and so on.

Using a remote hosting service, we basically let other people handle all of that.
In other words, we don’t need to worry for any of the hardware, software, and network connection issues—just the contents of our website.

There are numerous hosting services for static web pages.

For example, both **Google** and **Amazon**, as well as many other smaller companies, offer *paid* static hosting services.

We will use the **GitHub** platform for hosting our static web page hosting, which is *free*.

Although GitHub is mainly a platform for online storage of **Git** repositories and collaborative code development, one of its “side” functions is that of a static server.

The static server functionality of GitHub is known as **GitHub Pages**.
Using GitHub Pages as a remote static server has several advantages for our purposes:

- It is **simple**.
- It is **free**.
- It is part of **GitHub**, a popular platform for collaborative code development, which is useful to become familiar with.

Another good alternative is **surge.sh**. It is also free, and can be quicker to set up compared to GitHub Pages, but requires using the command line.

Git and GitHub

When working with code, it becomes important to keep track of different versions of your projects.

This allows you to undo changes made weeks or months ago.

Versioning becomes even more important when collaborating with others, since in that case you may need to split your project into several “branches”, or “merge” the changes contributed by several collaborators back together.

To do all of those things, people use **version-control systems**.

One of the most popular revision control systems around today is **Git.**

GitHub

Git is a version-control system, Git projects are also called repositories.

GitHub is a web-based Git repository hosting service. Basically, GitHub is an online service where you can store your Git repositories, either publicly or privately.

The platform also contains facilities for interacting with other people, such as raising and discussing issues or subscribing to updates on repositories and developers you are interested in, creating a community of online code-collaboration.

For anyone who wants to take part of open-source software development, using Git and GitHub is probably the most important skill after knowing how to write the code itself.

For any public GitHub repository, the user can trigger the **GitHub Pages** utility to serve the contents of the repository.

As a result, the contents of the repository will be automatically hosted at the following address:
https://GITHUB_USER_NAME.github.io/REPOSITORY_NAME/

where:

- GITHUB_USER_NAME is the **user name**
- REPOSITORY_NAME is the **repository name**

Everything about static servers applies in remote hosting too. The only difference is that the served directory is stored on another, remote server, rather than your own computer.

For example, in order for a web page to be loaded when one enters a repository URL as shown above, you need to have an index.html file in the root directory of your GitHub repository

Types of Website Hosting Services

Shared Hosting

A shared hosting service is suitable for small websites, blogs, and small businesses that are just starting out.

They are able to keep their costs down by allowing multiple websites to share the same server resources. This makes hosting your website affordable.
Suitable for: Starting a new blog or small business website.

Shared hosting provider: Bluehost

VPS Hosting

VPS hosting (Virtual Private Server hosting) is still a shared hosting environment. It offers a flexible set of resources to handle large traffic spikes.

You get a partitioned-off private server for your website that you can manage from your hosting control panel. This gives you the best of both worlds, the low cost of shared hosting with the flexibility of dedicated resources.

Suitable for: Medium-sized businesses, popular blogs, and eCommerce stores.

VPS hosting company: HostGator

Managed WordPress Hosting

Managed WordPress hosting is a specialized hosting service made specifically for WordPress.

On a managed hosting platform, the hosting company takes care of updates, backups, and caching of your website.

This allows you to focus on creating content and growing your business.

Suitable for: Popular blogs, business websites, membership websites.

Dedicated Hosting

A dedicated server hosting gives you the entire server dedicated to your own website.

You get all the resources of the server, advanced tools for server management, the ability to install your own software, and even your own operating system.

You'll be managing your own server, which may require some technical skills. It is an advanced option for larger websites that need high-performance to tackle higher traffic volume.

Suitable for: Enterprise-level businesses, hugely popular websites, eCommerce stores.

Dedicated hosting company: SiteGround or HostGator.

Questions

1. What is a web server? Kinds of Servers.
2. Can you explain the difference between an application server and a web server?
3. What is HTTP is?
4. Explain the Difference between URL and domain name
5. Web server software.

Test

1. There is an arrangement where you allow an agency to host your website for you on their Web Server for a fee, this is called as ?

- Web Service
- Web Hosting
- Web Marketing
- None of these

2. The World Wide Web is a massive collection of web sites, all hosted on

- computers
- Internet
- World Wide Web
- a network

3. The web server (computer) where your web site's html files, graphics, etc. reside is known as the

- web localhost
- web host
- web server
- None of these

4. WWW is the acronym of the

- Web World Wide
- World Wide Webpages
- World Wide Web
- World Wide Websites

The Web is a computer network all over the

- world wide
- country wide
- continent wide
- state wide

6. All the computers in the Web can communicate or transfer the data(text file) and graphics (picture file) with each other.

- False
- True
- Not always
- None of these

7. A Web browser access the web page from a web server by a

- request
- response
- Interrupts
- normal messages

8. A request is a standard HTTP request containing a

- computer address
- page address
- domain address
- MAC address

9. Which service is used to resolve the domain names to IP addresses ?

- Resolver
- Name-Converter
- DNS
- None of these

10. What is an ISP ?

- Internet Service Provider
- Internet State Provider
- Intranet Service Provider
- Internet State Provider

11. Which of these is not an expenses in the Web Service ?

- Hardware Expenses
- Software Expenses
- Labor Expenses
- Transprot Expenses

12. A domain name is the unique text name corresponding to the _______ of a computer on the Internet.

- numeric MAC address
- numeric network address
- numeric IP address
- All the above

13. What is used as a separator in a domain name ?

- dot(.)
- slash(/)
- colon(:)
- All the above

14. Users to your web site will often connect via a

- switch
- hub
- modem
- None of these