World Wide Web (WWW)

Mariana Kant 1

ORIGINS

1957

The Advanced Research Project Agency (**ARPA**) was created in response to the Soviet Union's successful launch of Sputnik, the world's first artificial satellite.

1962

ARPA started to work on some applications of computers to military technology. The Cold War raised concerns about the impact that a nuclear war could have on the integrity of computer networks to sustain military command and control.

A new concept was introduced: **distributed packet switching**. In this model network transmissions are split into small **packets** that may take different routes to their destination through different **nodes** (computers) along the network.

1969

The Department of Defense commissioned ARPANET for research into networking. First node was at UCLA followed by Stanford Research Institute, University of California at Santa Barbara, and University of Utah.

Mariana Kant 2

ORIGINS

Dr. Leonard Kleinrock is known as the inventor of Internet technology, having created the basic principles of packet switching while a graduate student at MIT.

1973

ARPANET linked 40 machines and had international connections to England and Norway.

One of the issues in computer communications is the reliability of messages send from one computer to another. Computers are different makes and models. They have different methods of sending and receiving packets of electronic information. There is also the issue of lost packets, when the information does not reach the intended computer because of transmission problems. These concerns led to the development of **TCP** (Transmission Control Protocol) to ensure reliable connections between diverse networks. Parallel development of **IP** (Internet Protocol) dealt with the problems of assembling packets of data and ensuring they reach the correct destination.

Mariana Kant

3

ORIGINS

1982

ARPANET use the TCP/IP protocol to communicate through a network of computers using land lines, radio, and satellite links.

At this moment, **"internet"** became defined as a set of networks interconnected through TCP/IP.

1984

1000 host computers were part of ARPANET, and Domain Name Servers (DNS) were introduced to permit use of host names (e.g., www.yorku.ca), in addition to numeric IP addresses (124.34.17.33), for identifying and linking computers on the networks.

Mariana Kant

4

ORIGINS

- On October 24, 1995, the Federal Networking Council unanimously passed a resolution defining the term Internet:

  "Internet" refers to the global information system that –
    - *(i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons;*
    - *(ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and*
    - *(iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein.*

Mariana Kant

ORIGINS

The Internet can be thought of as a technical infrastructure -- the computers, cables, networks, and switching mechanisms for one computer to communicate with another computer.

The benefits of networked computers are realized by the information being exchanged among people sitting behind the computers. Email and file transfer programs were, from the start, integral to the purpose behind the Internet, **putting people in touch with each other and with the information they need.**

1990

There were 300,000 computers connected by the network.

1995

Commercial use of the internet.1995-1997 the number of sites increased by over 6 million per year.

Today

There are approximately 604 million Internet users world wide.

Mariana Kant

ORIGINS

Although email and file transfer methods were important to Internet growth, they did not provide the "user friendly" methods needed by novice users to get to the growing repositories of information scattered around the world. It was still very much a technical issue to communicate through the Internet. Realization of the goal of an information superhighway required development of tools to "hide" Internet technology behind a human interface. This came with development of the World Wide Web and Internet browser software.

1989

Tim Berners-Lee, a consultant with the European Laboratory for Particle Physics (CERN), wrote a program entitled "Enquire-Within-Upon-Everything" which allowed links to be made between arbitrary text nodes in a document. Each node had a title identifier and a list of links so readers could jump from one section of a document to another by activating the text links.

ORIGINS

1990

Berners-Lee started work on a hypertext "browser."

He used the term "WorldWideWeb" as a name for the program and "World Wide Web" as a name for the project.

The Web included standard formats for text, graphics, sound, and video which could be indexed easily and searched by all networked machines.

Standards were proposed for:

- a Uniform Resource Locator (**URL**), which is the addressing scheme for the Web;
- the HyperText Transfer Protocol (**HTTP**), which is the set of network rules for transmitting Web pages;
- and the HyperText Markup Language (**HTML**), which is the language used to write the documents.

Format

The system uses hypertext: text with markup tags, and embedded links to text, images, audio files, video files in local or remote locations on the INTERNET.

<this is a tag>

The markup tags tell the Web browser how to display the page

The unit of transfer between two computers on the WWW is the document (page, resource).

HTML stands for Hyper Text Markup Language

XHTML stands for a reformulation of HTML as an XML application

An HTML/XHTML file is a text file containing markup tags

An HTML/XHTML file must have an .htm or .html file extension

INTERNET- WWW

The Internet is a global data communications system.
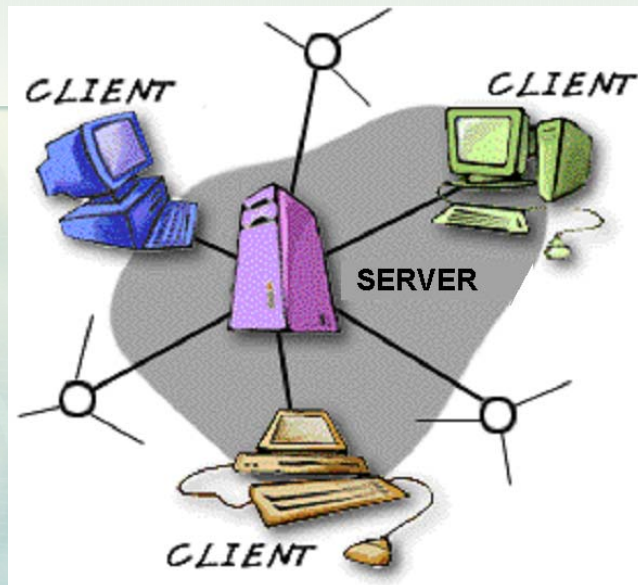
It is

- a hardware and
- software

infrastructure that provides connectivity between computers

The Web is one of the services communicated via the Internet.

It is a collection of interconnected documents and other resources, linked by hyperlinks and URLs.

The documents on the Web are stored on **servers** and we may access them using **Web clients** (**browsers**)

Web Servers

A **Web server** is a program that runs on a computer ℭ, connected to the INTERNET.

It is able:

- to find a document and send it to the requesting browser
- to run some program stored on ℭ (such programs dynamically generate documents that are returned to the browser)

There are different types of servers on the INTERNET. The difference comes from the protocol used to communicate.

HTTP, FTP, GOPHER, TELNET, MAILTO.

Web Servers (HTTP):

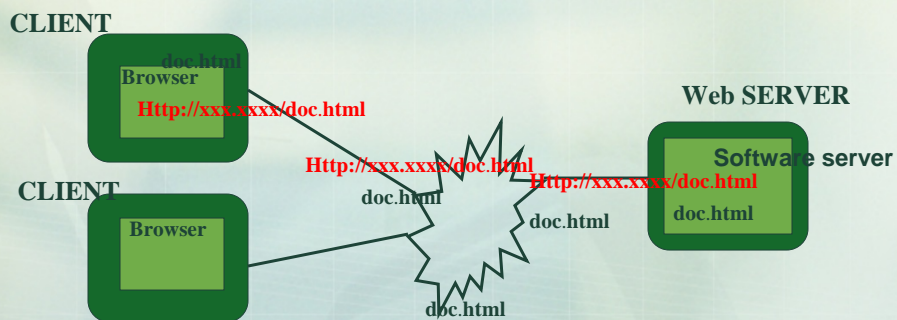APACHE, IIS (Internet Information System), WebSite Professional, etc.

Web Browsers

A Web client, browser, is a program that runs on a computer. If the computer is connected to the INTERNET then it is able:

- to send a query to a certain server concerning a document
- to display the document received from the server to the user using a GUI or a text interface.

Browser	Percent
Internet Explorer	78%
Firefox	16%
Chrome	3%
Safari	2%
Opera	1%

Source: Browser News

Mariana Kant

13

Client/SERVER model for HTTP protocol

CLIENT

doc.html

Browser

Http://xxx.xxxx/doc.html

Web SERVER

Software server

CLIENT

Http://xxx.xxxx/doc.html

doc.html

Http://xxx.xxxx/doc.html

doc.html

Browser

doc.html

doc.html

doc.html

Mariana Kant

14

Internet Protocol address- IP address

The individual computers in an organization are usually connected to each other in a LAN (Local Area Network). One node of the LAN is physically connected to the INTERNET.

INTERNET is a network of networks rather a network of computers.

A computer connected to the Internet has an IP address, four numbers separated by periods.

123.255.126.0.

0-255

Domain Names

Domain names offer a more intuitive way to name and find a website.

The domain names have different words, separated by periods, that indicate different levels of organization.

The last word in a domain is also called the *top-level domain*.

TLD	Meaning	Examples
.com	U.S. commercial business, a company	ibm.com, att.com, ford.com
.net	Network provider, Internet Service Provider	webtv.net
.gov	U.S. governmental agency	whitehouse.gov, nasa.gov
.edu	U.S. educational institution	uiuc.edu, stanford.edu
.org	Non-profit institution	redcross.org, sfopera.org
.mil	U.S. military	army.mil
.int	International	itu.int
.biz	Businesses	

Domain Names

Country domains. Each country in the world has its own top-level Internet domain with a unique alphabetic designation.

Austria	http://www.austria.at/
Australia	http://www.morning.com.au/
Barbados	http://www.barbados.gov.bb/
Bermuda	http://www.kbb.bm/
Brazil	http://home.iis.com.br/
Canada	http://www.canoe.ca/
Switzerland	http://www.badminton.ch
Chile	http://www.santiago.cl/
United States	http://www.co.maui.hi.us/

Domain Names

Search for a domain name
yorku.ca
cse.yorku.ca
All domain names are recorded in a central database, and each record in the database must be unique.

The database with the domain names resides on a **name server** and the software that access it is a **Domain Name System** (DNS).

A document request from a browser is routed to the nearest name server.

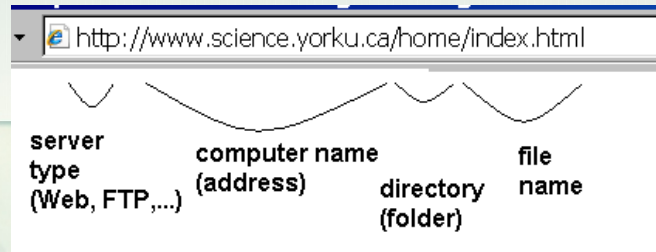The server converts the name in an IP address.

WEB HTTP clients
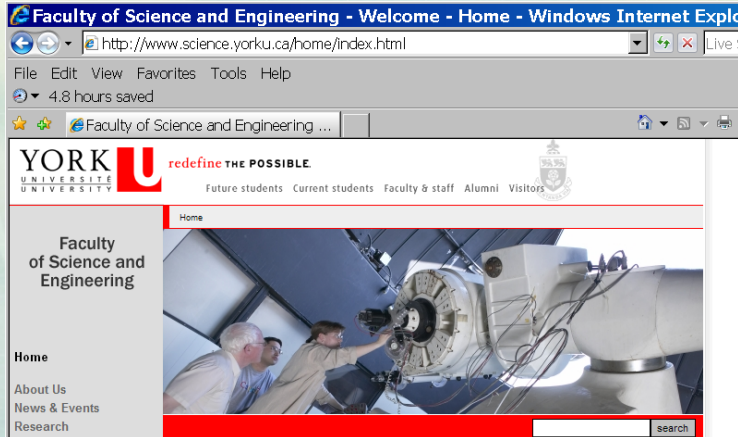
Universal Resource Locator

The URL is the address (the identification) used by the "client" to find a specific page on the WWW. We may see it as a phone number.
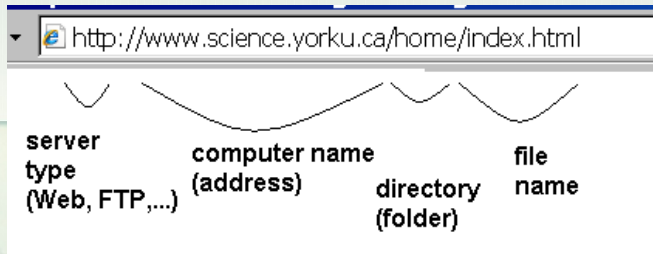
There are 3 components of this address:

- the service type (the type of server we will use)
- the address of the machine itself
- the path to the file we like to have

http://www.science.yorku.ca/home/index.html

server
type
(Web, FTP,...)

computer name
(address)

directory
(folder)

file
name

- the service type (using a specific communication protocol): http, ftp, gopher, telnet, etc.

- the address of the machine itself: it is the name of the server computer that runs the corresponding software which listens to a specific computer port (Web server :80)

- the path to the file we like to have (complete path, relative path)

FTP

File **T**ransfer **P**rotocol is an interactive file-transfer mechanism that enables you transfer files between computers that are connected to the Internet.
FTP was developed to allow transferring of files regardless of the operating system used (for example, between a Macintosh and Windows PC).

An **FTP client** is the software that you would use to connect to a remote

computer that is setup as an FTP server. The FTP client, such as **WinSCP**, sends a series of information to the server to "ask" the server if it can connect. If the server recognizes the information (e.g., your username and password), it will allow the connection.

An **FTP server** is a software program installed on a computer to allow or deny access to remote computers that attempt to access it. For example, a university might have installed **WinSCP's Server** so that professors and students can upload or download assignments and share files.

MIME: Multiple Internet Mail Extensions

MIME is the standard specification format for the documents on the WWW. The specifications apply to various kind of text, video data, or sound data. The MIME type has become the de facto standard for identifying content on the Internet.

In order to identify the type of attachment sent in an e-mail message, its MIME type is embedded in the message header.

A Web server attaches a MIME format specification to the beginning of the document it sends to the browser. The browser uses the format specification to determine which helper application or plug-in to launch.

MIME specifications has the form **type/subtype**.

Example: **text/plain, text/html, application/x-pdf** (adobe acrobat file), **image/gif**, **image/jpg, application/octet-stream (.exe file), audio/basic, audio/mpeg (mp3),** etc.

http://www.w3schools.com/media/media_mimeref.asp

Mariana Kant

◄ ►   23

MIME: Multiple Internet Mail Extensions

Web servers determine the type of a document by using the file's name extension as the key for the key into **its table of types**.

Example: The extension .html tells to the server that it should attach *text/html* to the document before sending it to the requesting browser.

Experimental subtype

An experimental subtype name starts with **x-** : video/x-wav (wav files).

When a Web provider adds an experimental subtype, he/she must provide the program/module able to read and execute the corresponding file.

■ If the program is external to the browser we call it **a helper application**. The browser will call it to be able to display/execute the data in the file. (?.doc)

■ If it is a module (or many) which are inserted in the browser (Firefox updates) we call it **a plug-in**. The browser will execute it to display the file.
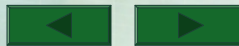
Mariana Kant

◄ ►   24

MIME: Multiple Internet Mail Extensions

The browser determines the helper application (MS-Word) or plug-in it

needs by examining the browser configuration file, which provides an

association between file types and their required helpers or plug-ins.

```
<head>

<meta http-equiv=Content-Type content="text/html;
   charset=windows-1252">

</head>
```

XHTML
EXtensible Hypertext Markup Language

XHTML is a reformulation of HTML as an XML application

The hypertext files are written using the HTML/XHTML language.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
   "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
   <head>
      <title>2010_Float </title></head>
<body link="#0000ff" vlink="#800080" alink="#ff0000">
<img src = "ble.gif" style = "float: right" />
This is the image we like to see on the right and the text on the left
</body>
</html>
```

The html/xhtml tags may specify also other html/xhtml files, image files (*.gif,
   *.jpg ...), audio and video files.

EXtensible HTML versus Cascading Style Sheets

A web document has two aspects: *content* and *styling*.
The *content* consists of the words, possibly accompanied by images and other elements, that are necessary to convey the information and the meaning of our web page.
Styling refers to the graphic or artistic design of the web page, such as the fonts, colors, or background images that we apply to a web page to make it look appealing. We can change the fonts or colors without changing the meaning or information content of your web page.

In modern, high-quality web design, we separate the content (XHTML) from the styling (CSS) so that we will be able to easily change the graphic design without messing up the content, or change the content without destroying the graphic design.

Mariana Kant

◄ ►

27

Good luck!

Mariana Kant

28