ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Introduction to Web Technologies and e-Services

Lecture contents

1. Internet, Web
2. HTTP
3. URL
4. Web Browser
5. Web Application
6. Web Application Architecture
7. Web Developer Roadmap

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Reasonable Questions

- What is the World Wide Web?

- Is it the same thing as the Internet?

- Who invented it?

- How old is it?

- How does it work?

- What kinds of things can it do?

- What does it have to do with programming?

Web ≠ Internet

- Internet: a physical network connecting millions of computers using the same protocols for sharing/transmitting information (TCP/IP)
  - in reality, the Internet is a network of smaller networks

- World Wide Web: a collection of interlinked multimedia documents that are stored on the Internet and accessed using a common protocol (HTTP)
- Key distinction: Internet is hardware; Web is software along with data, documents, and other media

- Many other Internet-based applications exist e.g., email, telnet, ftp, usenet, instant messaging services, file-sharing services, …

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

(A Very Brief) History of the Internet

- The idea of a long-distance computer network traces back to early 60's

    - Joseph Licklider at M.I.T. (a "time-sharing network of computers")

    - Paul Baran at Rand   (tasked with designing a "survivable" communications system that could maintain communication between end points even after damage from a nuclear attack)

    - Donald Davies at National Physics Laboratory in U.K.

- In particular, the US Department of Defense was interested in the development of distributed, decentralized networks

    - survivability (i.e., network still functions despite a local attack)

    - fault-tolerance (i.e., network still functions despite local failure)

    - *contrast with phone system, electrical system which are highly centralized services*
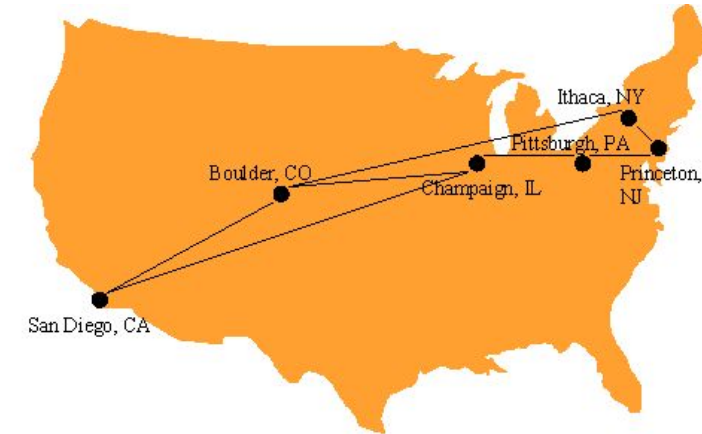
The Internet

- In 1969, Advanced Research Project Agency funded the ARPANET
  - Connected computers at UC Los Angeles, UC Santa Barbara, Stanford Research Institute, and University of Utah
  - Allowed researchers to share data, communicate
    - 56Kb/sec communication lines (vs. 110 b/sec over phone lines)
- Technical origin
  - One of earliest attempts to network heterogeneous, geographically dispersed computers
  - Email first available on ARPANET in 1972 (and quickly very popular!)

The Internet

The 6 supercomputer centers connected by the early NSFNET backbone

- Open-access networks
  - Regional university networks (e.g., SURAnet)
  - CSNET for CS departments not on ARPANET
- NSFNET (1985-1995)
  - Primary purpose: connect supercomputer centers
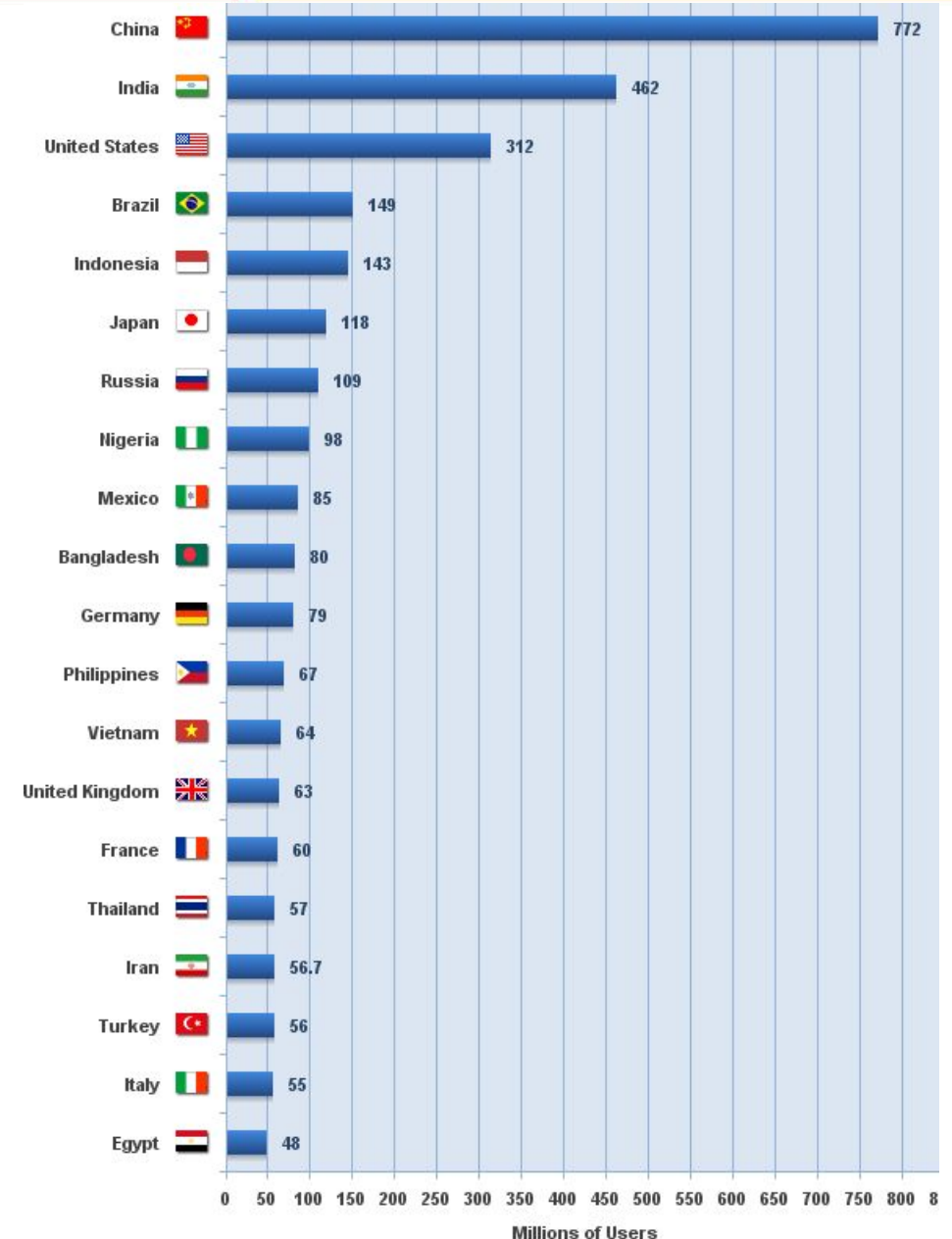 - Secondary purpose: provide backbone to connect regional networks

Internet Growth

- Throughout the 70's, the size of the ARPANET doubled every year
  - First ARPANET e-mail sent in 1971
  - Decentralization mades adding new computers easy
  - TCP/IP developed in the mid 1970s for more efficient packet routing
  - Migration of ARPANET to TCP/IP completed 1 january, 1983
  - ~1000 military & academic host computers connected by 1984
- In 80's, U.S. government took a larger role in Internet development
  - Created NSFNET for academic research in 1986
  - ARPANET was retained for military & government computers
- By 90's, Internet connected virtually all colleges & universities
  - Businesses and individuals also connecting as computing costs fell
  - ~1,000,000 computers by 1992
- In 1992, control of the Internet was transferred to a non-profit organizations
  - Internet society: Internet engineering task force, Internet architecture board, Internet assigned number authority, World-wide-web consortium (W3C)

Internet Growth (cont.)

- Internet has exhibited exponential growth, doubling in size every 1-2 years   (stats from *Internet Software Consortium)*

World Top 20 Countries in Internet Consumption
at December 31, 2018

| Country	Millions of Users
China	772
India	462
United States	312
Brazil	149
Indonesia	143
Japan	118
Russia	109
Nigeria	98
Mexico	85
Bangladesh	80
Germany	79
Philippines	67
Vietnam	64
United Kingdom	63
France	60
Thailand	57
Iran	56.7
Turkey	56
Italy	55
Egypt	48

(A Very Brief) History of the Web

- The idea of hypertext (cross-linked and inter-linked documents) traces back to Vannevar Bush in the 1940's
  - Online hypertext systems began to be developed in 1960's
    - e.g., Ted Nelson and Andy Van Dam's hypertext editing system (HES), Doug Englebert's NLS (on-line system)
  - In 1987, Apple introduced hypercard (a hypermedia system that predated the WWW)
- In 1989, Tim Berners-lee at the European particle physics laboratory (CERN) designed a hypertext system for linking documents over the internet
  - Designed a (non-wysiwyg) language for specifying document content
    - Evolved into hypertext markup language (HTML)
  - Designed a protocol for downloading documents and interpreting the content
    - Evolved into hypertext transfer protocol (HTTP)
  - Implemented the first browser -- text-based, no embedded media

The web was born!

History of the Web (cont.)

- The Web was an obscure, European research tool until 1993

- In 1993, Marc Andreessen and Eric Bina (at the National Center for Supercomputing Applications, a unit of the University of Illinois) developed Mosaic, one of the early graphical Web browsers that popularized the WWW for the general public (Erwise was the first one, ViolaWWW the second)
  - The intuitive, clickable interface helped make hypertext accessible to the masses
  - Made the integration of multimedia (images, video, sound, …) much easier
  - Andreessen left NCSA to found Netscape in 1994
    - Cheap/free browser further popularized the Web (75% market share in 1996)

- In 1995, Microsoft came out with Internet Explorer

- Opera web browser released in 1996
  - Netscape bought by AOL in 1998 for US$4.2 billion in stock
  - Firefox web browser, version 1.0, released in 2004
  - Google Chrome released in 2008

- Today, the Web is the most visible aspect of the Internet

History of the Web (cont.)
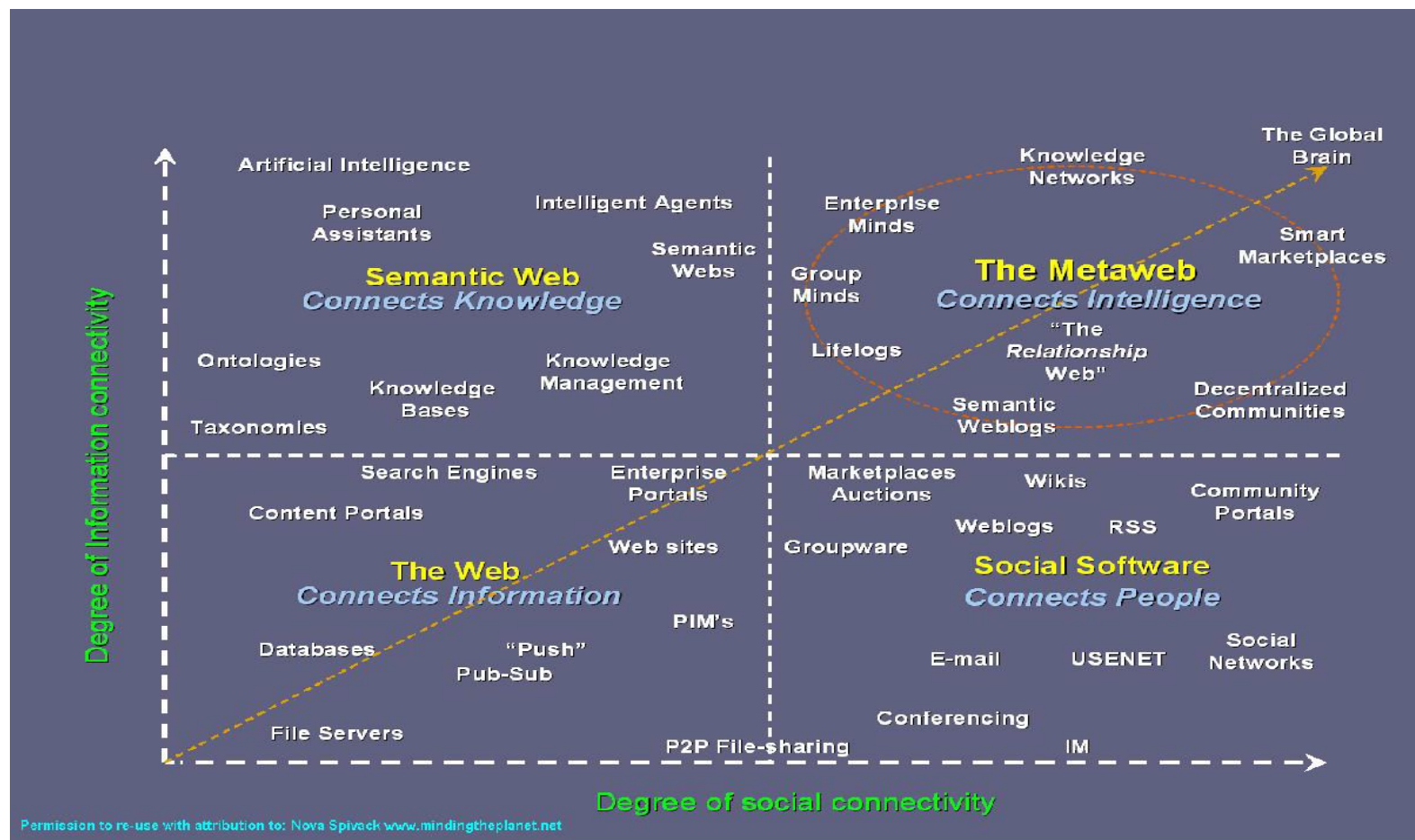
World Wide Web

- **The Web** is the collection of machines (Web servers) on the Internet that provide information, particularly HTML documents, via HTTP.

- Machines that access information on the Web are known as **Web clients**. A **Web browser** is software used by an end user to access the Web.

Hypertext Transport Protocol (HTTP)

- HTTP is based on the request-response communication model:
  - Client sends a request
  - Server sends a response

- HTTP is a stateless protocol:
  - The protocol does not require the server to remember anything about the client between requests.

HTTP

- Normally implemented over a TCP connection (80 is standard port number for HTTP)
- Typical browser-server interaction:
  - User enters Web address in browser
  - Browser uses DNS to locate IP address
  - Browser opens TCP connection to server
  - Browser sends HTTP request over connection
  - Server sends HTTP response to browser over connection
  - Browser displays body of response in the client area of the browser window

HTTP Request

- Structure of the request:
  - start line
  - header field(s)
  - blank line
  - optional body

HTTP Request

- Structure of the request:
  - **start line**
  - header field(s)
  - blank line
  - optional body

HTTP Request

- Start line
  - Example: GET / HTTP/1.1
- Three space-separated parts:
  - HTTP request method
  - Request-URI (Uniform Resource Identifier)
  - HTTP version

HTTP Request

- Start line
  - Example: GET / HTTP/1.1
- Three space-separated parts:
  - HTTP request method
  - Request-URI
  - **HTTP version**
    - We will cover 1.1, in which version part of start line must be exactly as shown

HTTP Request

- Start line
  - Example: GET / HTTP/1.1
- Three space-separated parts:
  - HTTP request method
  - **Request-URI**
  - HTTP version

HTTP Request

- Uniform Resource Identifier (URI)
  - Syntax: *scheme* : *scheme-depend-part*

    - Ex: In `http://www.example.com/`

    - the scheme is `http`

  - Request-URI is the portion of the requested URI that follows the host name (which is supplied by the required Host header field)
    - Ex: `/` is Request-URI portion of `http://www.example.com/`

URI

- URI's are of two types:
  - Uniform Resource Name (URN)
    - o Can be used to identify resources with unique names, such as books (which have unique ISBN's)
    - o Scheme is urn

  - Uniform Resource Locator (URL)
    - o Specifies location at which a resource can be found
    - o In addition to http, some other URL schemes are https, ftp, mailto, and file

HTTP Response

- Structure of the response:
  - status line
  - header field(s)
  - blank line
  - optional body

HTTP Response

- Structure of the response:
  - **status line**
  - header field(s)
  - blank line
  - optional body

HTTP Response

- Status line
  - Example: HTTP/1.1 200 OK
- Three space-separated parts:
  - HTTP version
  - status code
  - reason phrase (intended for human use)

HTTP Response

- Status code
  - Three-digit number
  - First digit is class of the status code:
    - 1=Informational
    - 2=Success
    - 3=Redirection (alternate URL is supplied)
    - 4=Client Error
    - 5=Server Error
  - Other two digits provide additional information
  - See http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html

HTTP Response

- Structure of the response:
  - status line
  - **header field(s)**
  - blank line
  - optional body

HTTP Response

- Common header fields:
  - Connection, Content-Type, Content-Length
  - Date: date and time at which response was generated (required)
  - Location: alternate URI if status is redirection
  - Last-Modified: date and time the requested resource was last modified on the server
  - Expires: date and time after which the client's copy of the resource will be out-of-date
  - ETag: a unique identifier for this version of the requested resource (changes if resource changes)

HTTP Request/Response Examples

Connect { $ **telnet www.example.org 80**
Trying 192.0.34.166...
Connected to www.example.com
(192.0.34.166).
Escape character is '^]'.

Send
Request { **GET / HTTP/1.1**
Host: www.example.org

Receive
Response { HTTP/1.1 200 OK
Date: Thu, 09 Oct 2003 20:30:49
GMT
…

Web Growth

How Many Websites Are There?
Number of websites online from 1991 to 2019
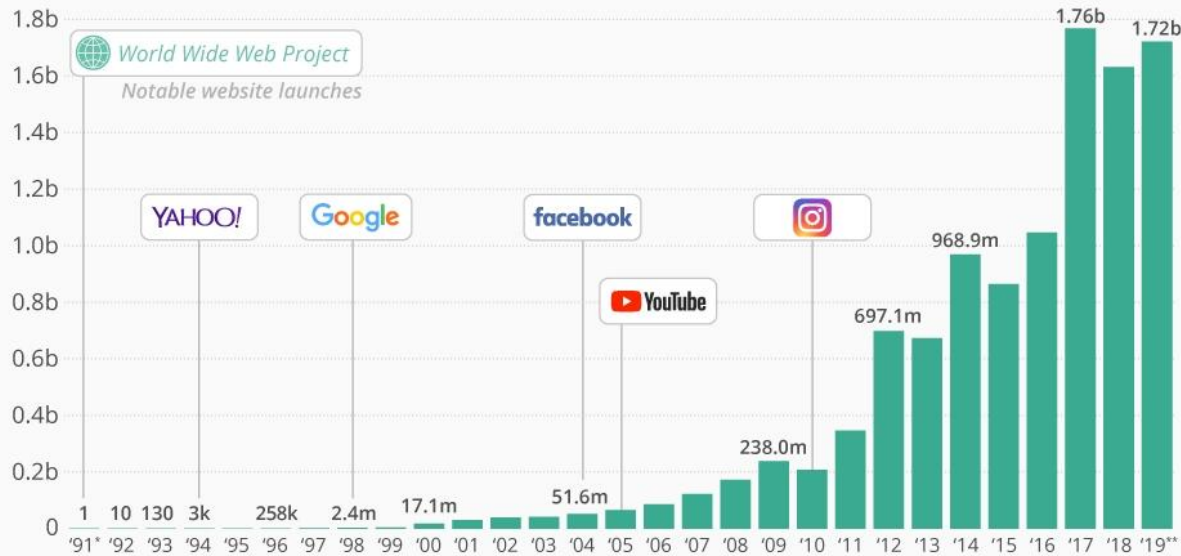
Year	Computers on the Internet	Web Servers on the Internet
2011	~605,000,000	~250,000,000
2008		172,338,726
2006	439,286,364	85,541,228
2004	285,139,107	56,923,737
2002	162,128,493	33,082,657
2000	93,047,785	18,169,498
1998	36,739,000	4,279,000
1996	12,881,000	300,000
1994	3,212,000	3,000
1992	992,000	50

Google Chrome → 2008
Firefox → 2004
Safari → 2002
IE, Opera → 1994
Netscape → 1994
Mosaic → 1992

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Web growth (cont.)

- Internet addresses are used to identify computers on the internet.

- Internet Protocol version 4 (IPv4) was first defined in 1981 and is still in use today, but this uses a 32-bit number to specify addresses.

- IPv4 provides around 4.29 billion addresses that are in use (or reserved).

- The Internet Assigned Numbers Authority exhausted their allocated addresses on January 31, 2011, and the Asia-Pacific Network Information Centre (regional internet registry) exhausted theirs on April 15, 2011. Also, the Réseaux IP Européens Network Coordination Centre (regional internet registry for Europe, the Middle East, and parts of central Asia) exhausted their allotment on 14 September, 2012. Other regional internet registries are predicted to run out within a few years.

- IPv6 had been deployed since the mid-2000s and uses 128 bit addresses, but also redesigned to allow more efficient routing, network aggregation, and ease of network reconfiguration.

Web Browsers

- First graphical browser running on general-purpose platforms:

Web Browsers
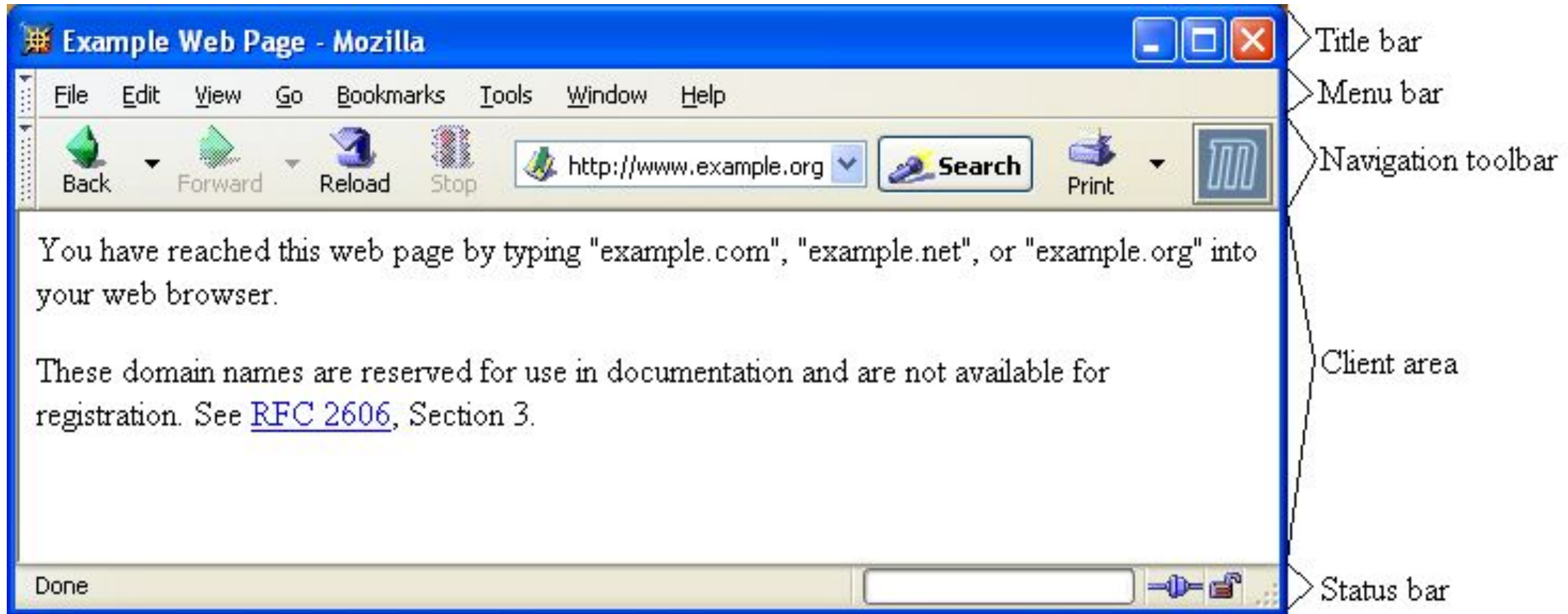
Web Browsers

- Primary tasks:
  - Convert web addresses (URL's) to HTTP requests
  - Communicate with web servers via HTTP
  - Render (appropriately display) documents returned by a server

Static vs. Dynamic pages

- Most Web pages are *static*
  - contents (text/links/images) are the same each time it is accessed

    e.g., online documents, most homepages

    HyperText Markup Language (HTML) is used to specify text/image format
- As the Web continues to move towards more and more online services and e-commerce continues to grow, Web pages must also provide *dynamic* content
  - pages can be fluid, changeable (e.g., rotating banners)
  - must be able to react to the user's actions, request and process info, tailor services

    e.g., amazon.com
- This course is about applying your programming skills to the development of dynamic Web pages and applications

Web Clients

- Many possible web clients:
  - Text-only "browser" (lynx)
  - Mobile phones
  - Robots (software-only clients, e.g., search engine "crawlers")
  - etc.

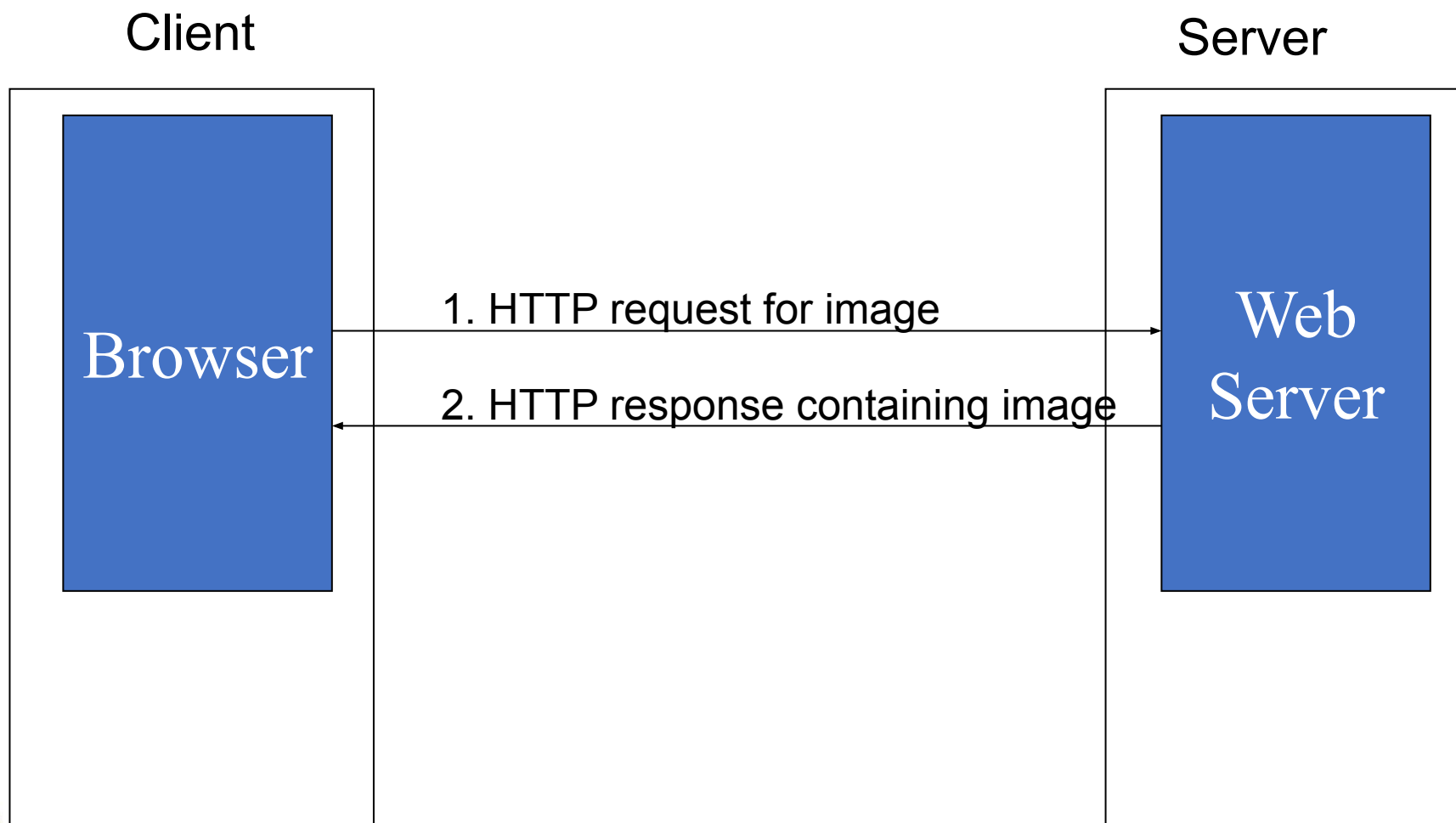# Web Servers

- Basic functionality:
  - Receive HTTP request via TCP
  - Map host header (domain name) to specific virtual host (one of many host names sharing an IP address)
  - Map Request-URI to specific resource associated with the virtual host
    - File: Return file in HTTP response
    - Program: Run program and return output in HTTP response
  - Map type of resource to appropriate MIME type and use to set Content-Type header in HTTP response
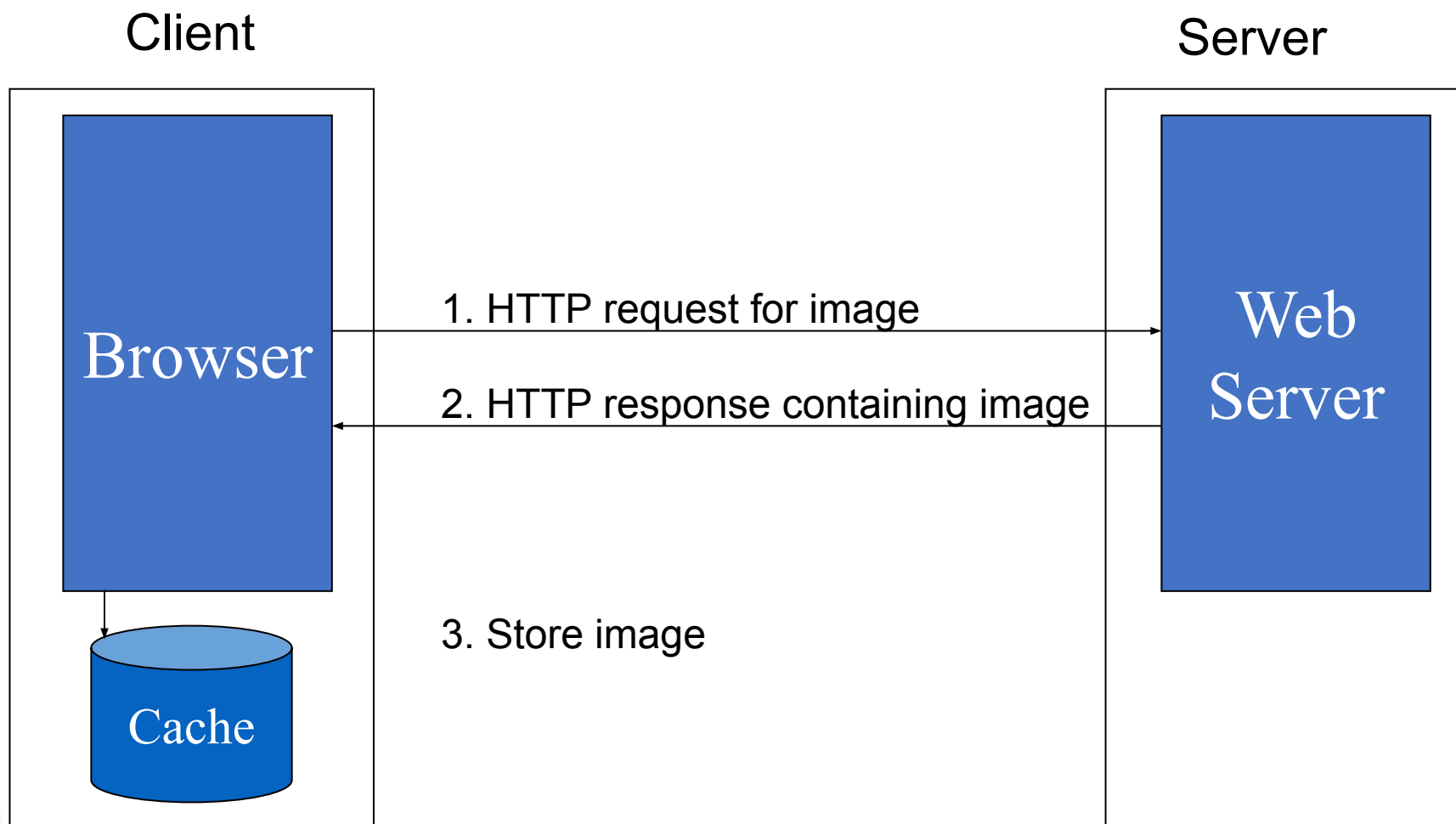 - Log information about the request and response

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Web Servers

- httpd: UIUC, primary Web server c. 1995
- Apache: "A patchy" version of httpd, now the most popular server (esp. on Linux platforms)
- IIS: Microsoft Internet Information Server
- Tomcat:
  - Java-based
  - Provides container (Catalina) for running Java servlets (HTML-generating programs) as back-end to Apache or IIS
  - Can run stand-alone using Coyote HTTP front-end
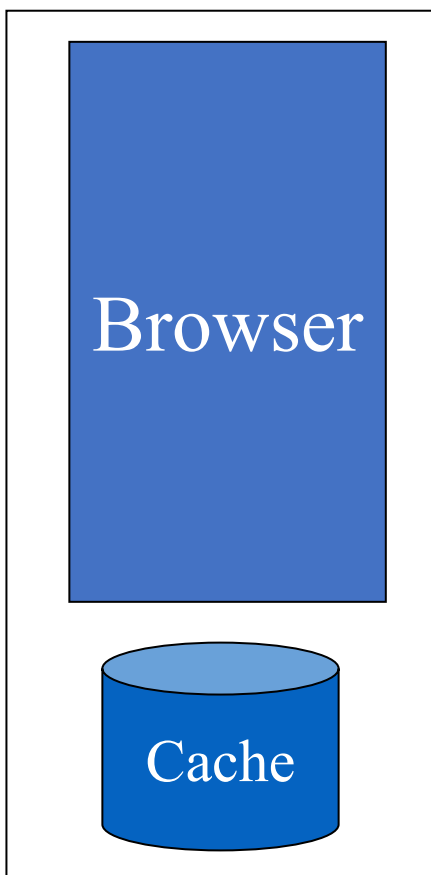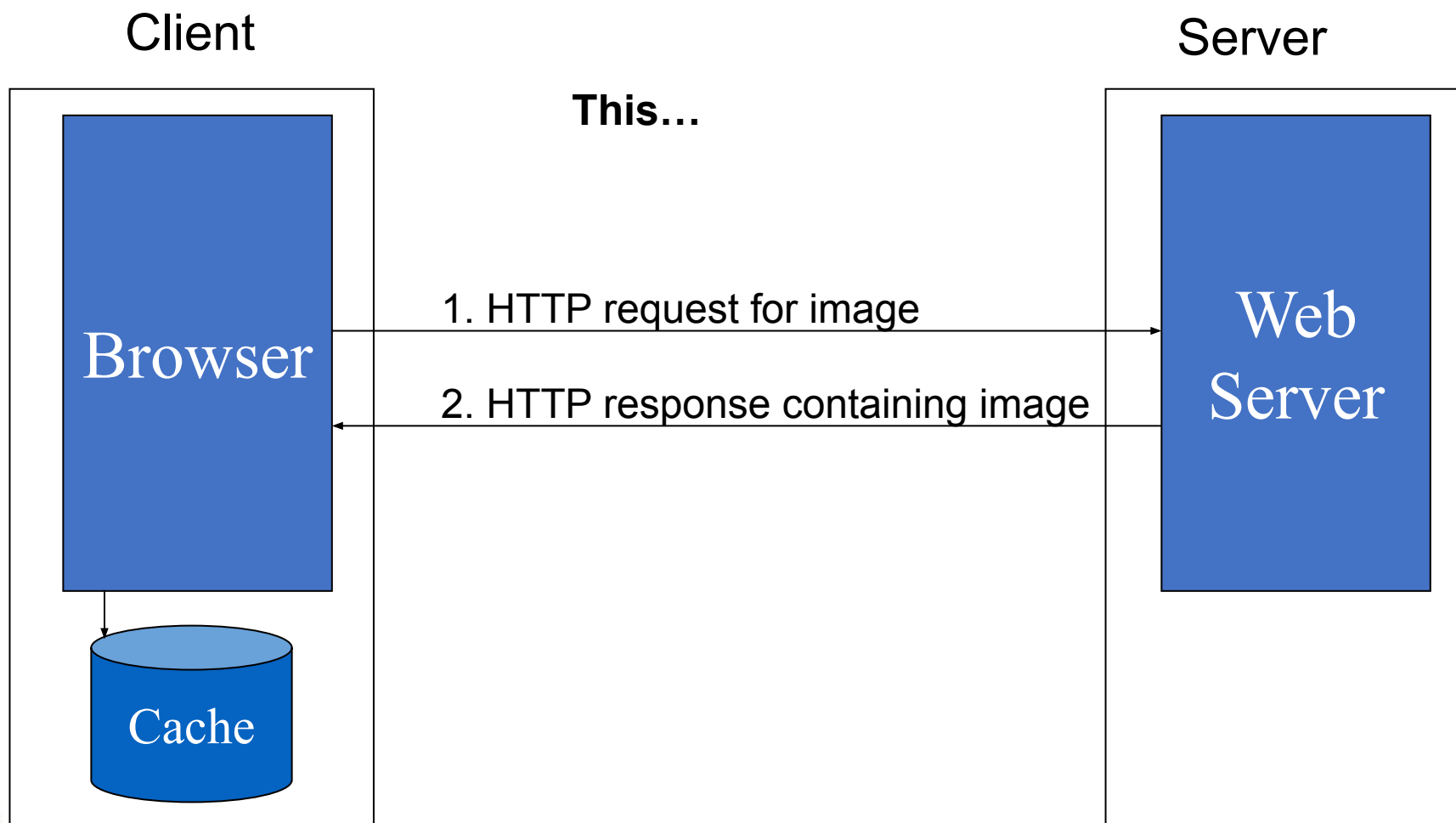
Web server/client

Web server/client

Client

Server

Browser

Web
Server

1. HTTP request for image

2. HTTP response containing image

3. Store image

Cache

Web server/client

Client

Server

Browser

I need that
image again

Web
Server

Cache

Web server/client

Client

Server

This…

Browser

Web Server

1. HTTP request for image

2. HTTP response containing image

Cache

Web server/client

Client

Server

Browser

Web Server

Get image

… or this

Cache

Web Developer Roadmap 2020

Required for any path

- Git - Version Control
- Basic Terminal Usage
- Data Structures & Algorithms
- GitHub
- Licenses
- Semantic Versioning
- SSH
- HTTP/HTTPS and APIs
- Design Patterns
- Character Encodings

Web Developer in 2020

Choose your path
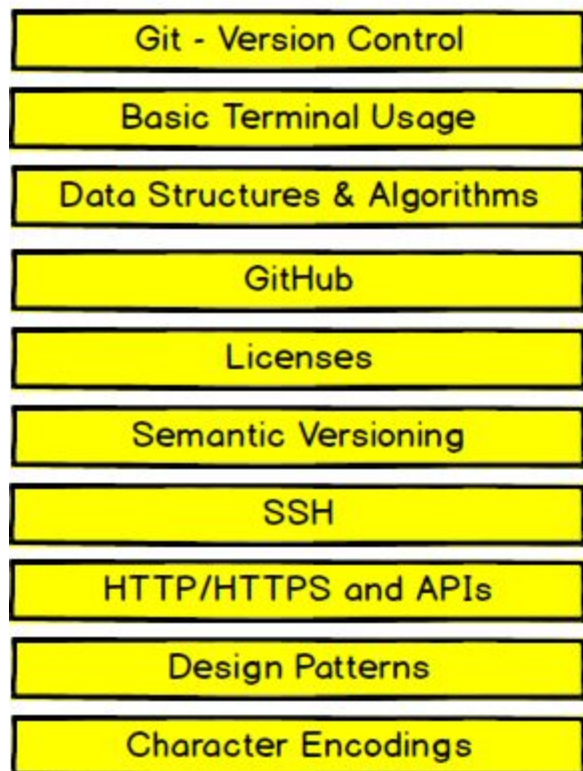
- Front-end
- Back-end
- DevOps
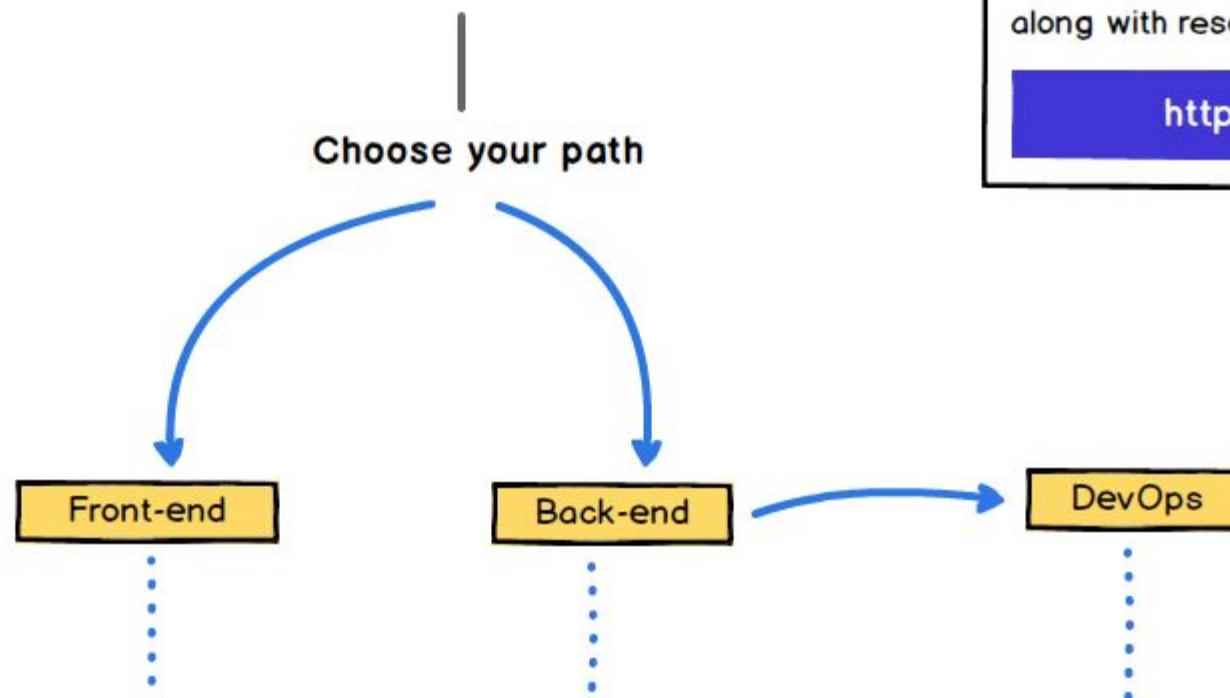
Find the detailed version of this roadmap along with resources and other roadmaps

http:// roadmap.sh

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Front-end

Back-end

DevOps

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG