CT219H – Web Programming Fundamentals

Chapter 1
Internet, WWW and HTML

2021

Tran Cong An
(tcan@cit.ctu.edu.vn)

Content

- Internet
- World Wide Web (WWW)
- HyperText Markup Language (HTML)

The Internet

What the Internet is?

- A world-wide network of computer networks (a network of networks)
- At the lowest level, since 1982, all connections use TCP/IP
- TCP/IP hides the differences among devices connected to the Internet

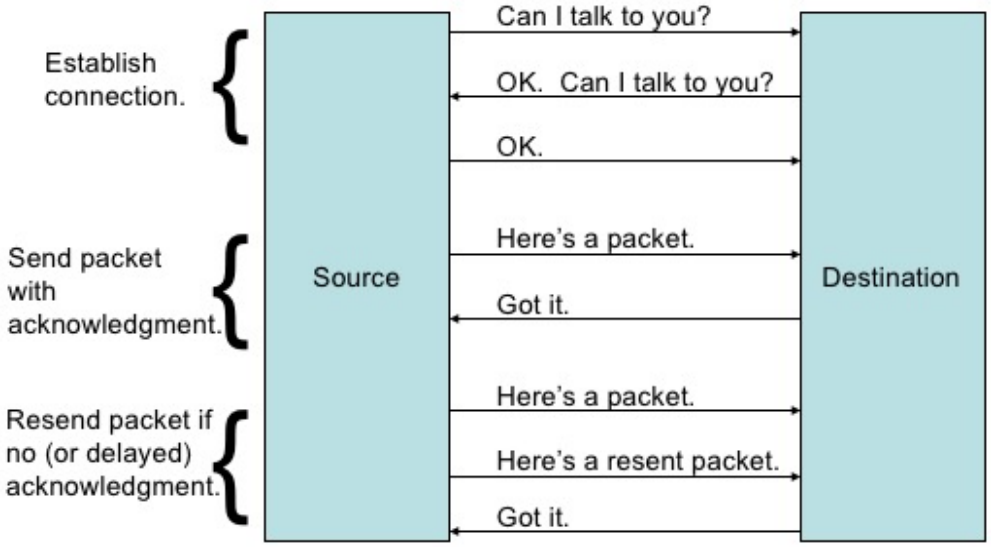
- Origin of the Internet:
  - ARPAnet - late 1960s and early 1970s (for military uses)
  - Network reliability
  - For ARPA-funded research organizations

The Internet basic concepts

- **Communication protocol**: how computers talk?
  - telephone “protocol”: how you answer and end call, what language you speak, etc
- **Internet Protocol** (IP):
  - A protocol to address computers in Internet (every node has a unique numeric address)
  - 32-bit number (in IPv4) or 128-bit number (in IPv6)
  - Associated with at most one device at a time (although device may have more than one)
  - Written as four dot-separated bytes, e.g. 192.168.1.123 (IPv4)

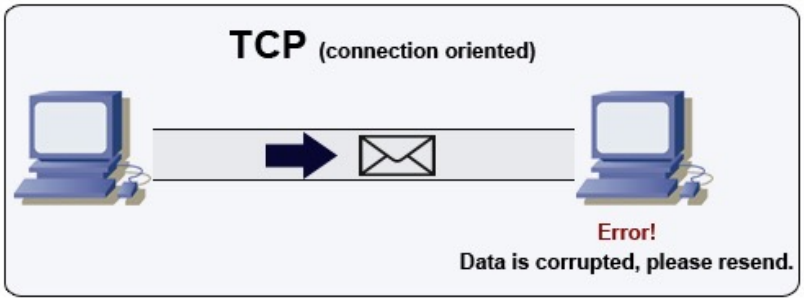
The Internet basic concepts

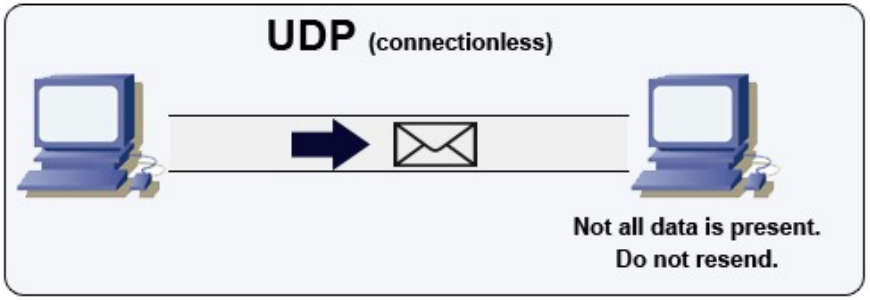
- **TCP** (Transmission Control Protocol): adds concept of a connection and port on top of IP
  - Provides guarantee that packets delivered
  - Provide two-way (full duplex) communication



The Internet basic concepts

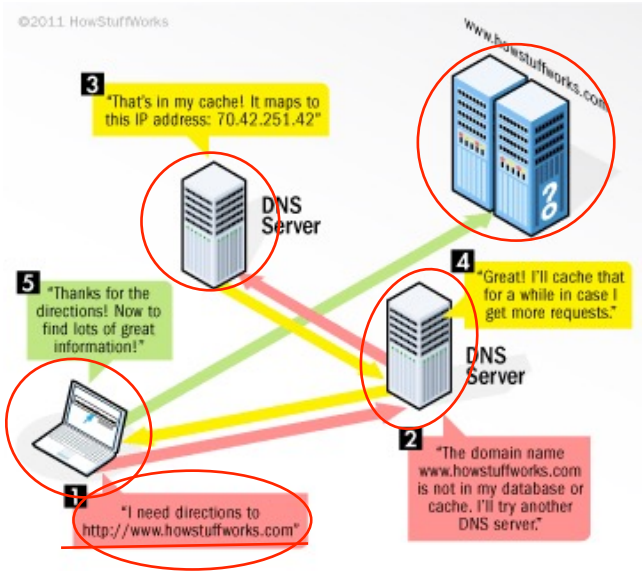
- **UDP** (User Datagram Protocol):
  - Builds on IP and provides port concept
  - No connection concept
  - No transmission guarantee
  - Lightweight, so faster for one-time messages





The Internet basic concepts

- **DNS** (Domain name system):
  - Form: host.domain, for example: www.ctu.edu.vn
  - First domain is the smallest, last is the largest (top level domain)
  - Last domain specifies the type (or location) of organization
  - DNS servers: convert DNS to IP



Analogy to Telephone Network

- IP ~ the telephone network
- TCP ~ calling someone who answers, having a conversation, and hanging up
- UDP ~ calling someone and leaving a message
- DNS ~ directory assistance

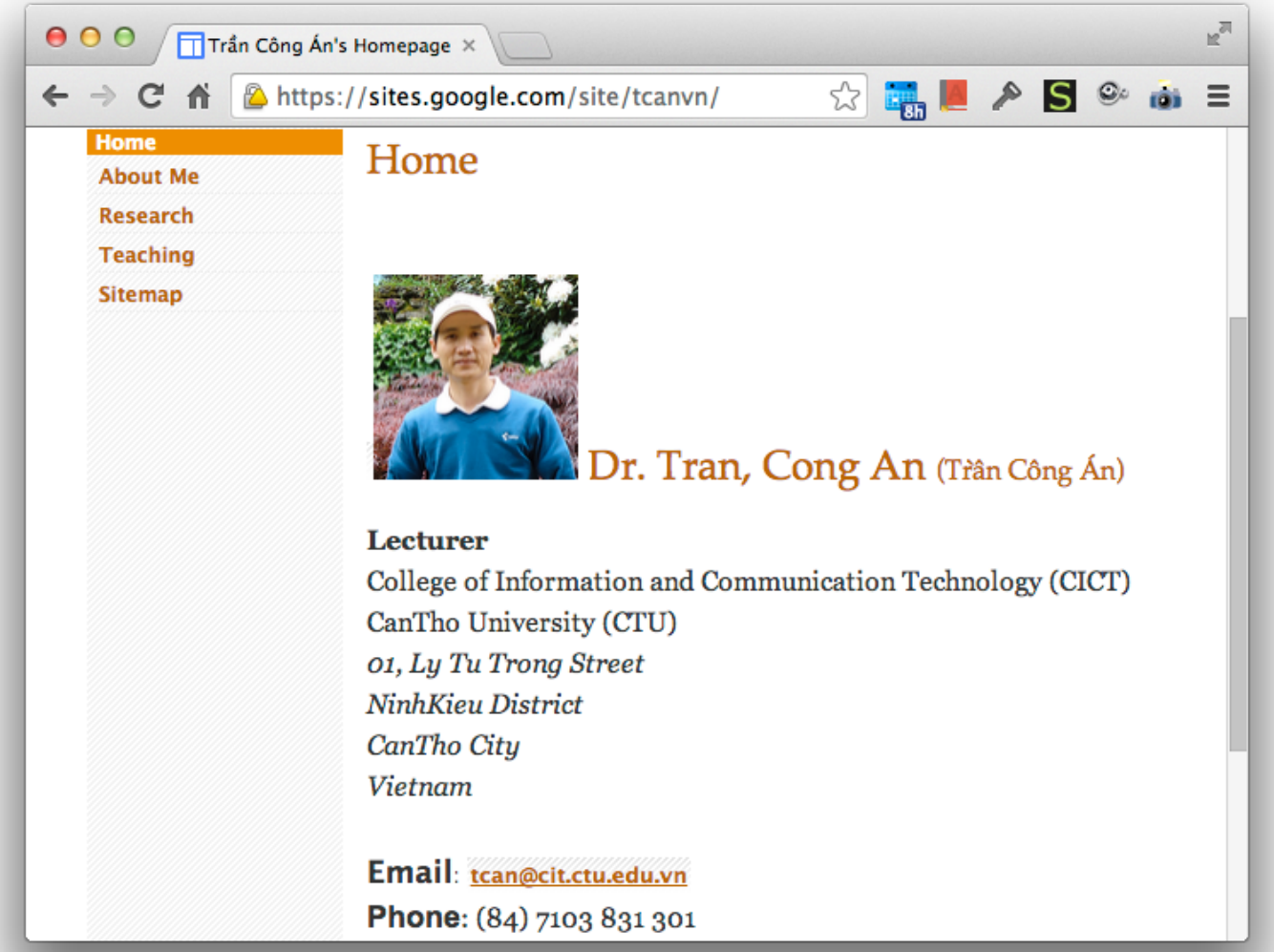
The WWW

What is WWW?

- One of services for organizing Internet-based information
  - Document form: HyperText
  - Information representation using HTML (HyperText Markup Language)
  - Communication protocol: HTTP (HyperText Transfer Protocol)

- Origins:
  - Tim Berners-Lee at CERN (European Organization for Nuclear Research) in 1989
  - Purpose: to allow scientists to have access to many databases of scientific work (share) through their own computers

WWW Document

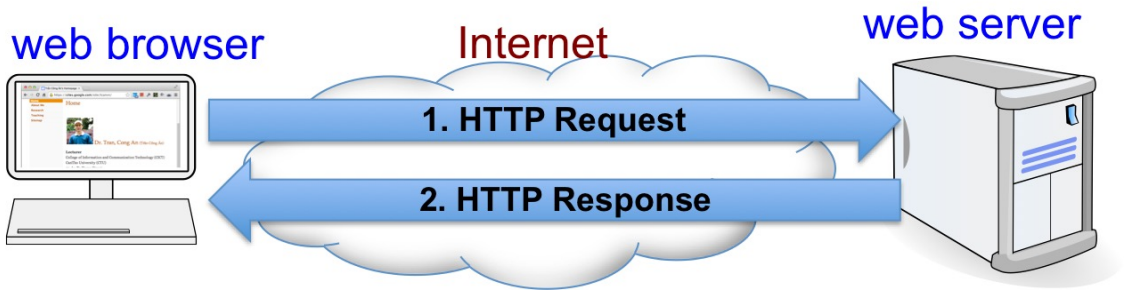


Hypertext, Webpage and Website

- HyperText: text displayed on a computer display, which contains links to other texts (hyperlinks)
- Webpage:
  - a document commonly written in HTML
  - accessible through the Internet or other networks using an Internet browser
- Website:
  - a collection of related webpages
  - typically identified with a common domain name, and published on at least one web server.

Web Server and Web Browser

- The web is a collection of Web Servers on the Internet that provides HTTP documents via HTTP protocol, either existing documents (static web) or dynamically built documents (dynamic web)
  - Some popular web servers: Apache (open source), IIS (Windows)
- A Web Browser is a computer program that is used by end users to access the Web
  - Popular web browsers: Chrome, Firefox, Microsoft Edge,…



Web Server and Web Browser

- Additional functionality of Web Browsers:
  - Execution of scripts (e.g., drop-down menus)
  - Event handling (e.g., mouse clicks)
  - GUI for controls (e.g., buttons)
  - Secure communication with servers
  - Display of non-HTML documents (e.g., PDF) via plug-ins

Web Browser and Web Server

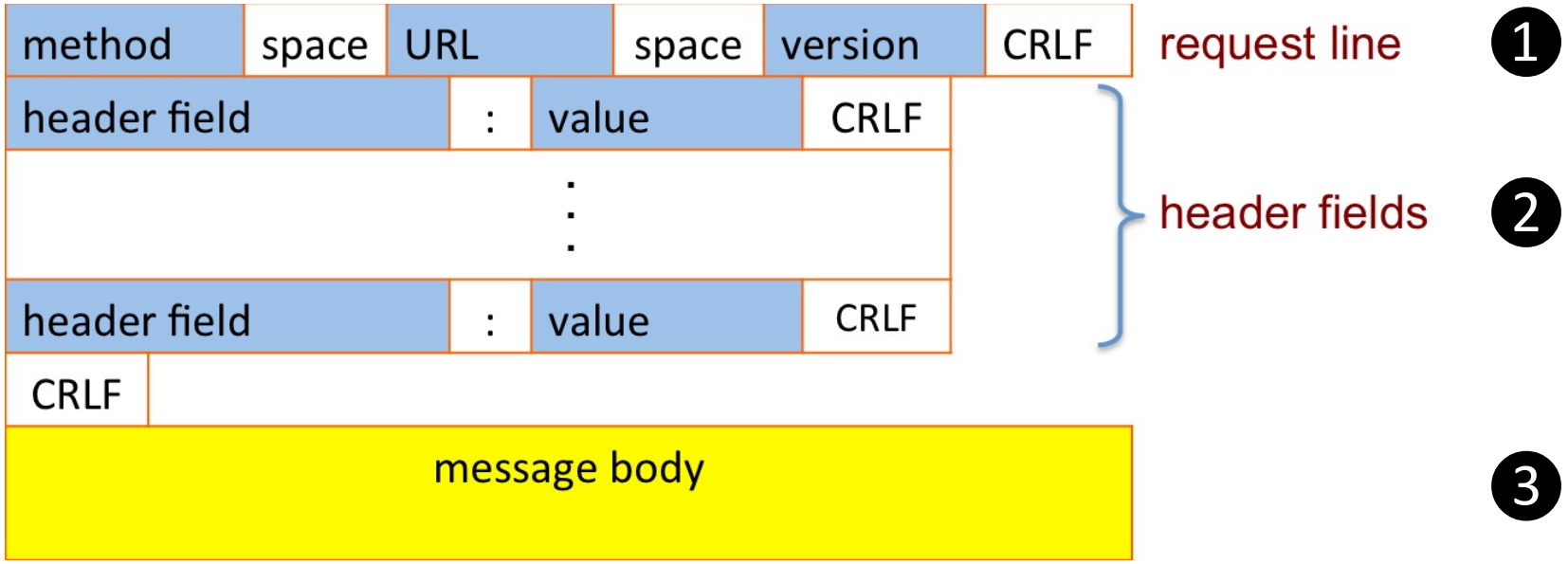
- Typical browser-server interaction:
  1) User enters Web address in browser (DNS)
  2) Browser uses DNS to locate IP address
  3) Browser opens TCP connection to server
  4) Browser sends HTTP request over connection
  5) Server sends HTTP response to browser over connection
  6) Browser displays body of response in the client area of the browser window

HyperText Transfer Protocol (HTTP)

- Communication protocol between browsers and web servers
  - A set of rules that are used in communication between web browsers and web servers
- HTTP is a stateless protocol:
  - The protocol does not require the server to remember anything about the client between requests
- Normally implemented over TCP protocol
- Information transmitted: HyperText (plant text)
- HTTP versions: 0.9, 1.0, 1.1 (currently)

HyperText Transfer Protocol (HTTP)

Structure of a HTTP request package



HyperText Transfer Protocol (HTTP)

1) Request line:
   - method: GET, POST, HEAD, PUT, DELETE,…
   - URL: path to the requested resource
   - version: HTTP version used in communication between client and server (currently 1.1)
   - Example: **`GET / HTTP/1.1`**
     (request the homepage using HTTP protocol version 1.1)
2) Header fields: additional information with the request
3) Body: usually empty, except in case browser wants to upload

HyperText Transfer Protocol (HTTP)

- Request methods (case sensitive):
  - **GET**: request a resource from the server given a URL
  - **POST**: used to send data to server (using HTML forms)
  - **HEAD**: same as GET method but get the status and header section only

  - PUT: replaces resources identified by a URL with the uploaded content
  - DELETE: removes resources given by a URL

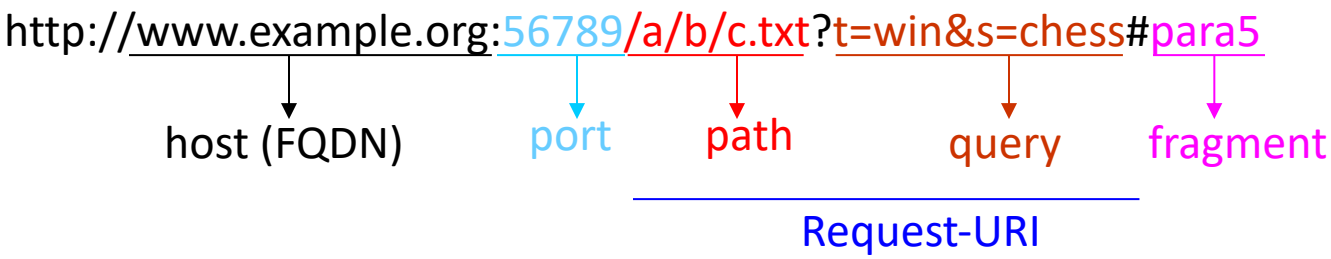
Note: Not all request methods are allowed by a web server

HyperText Transfer Protocol (HTTP)

- URL:
  - Specifies location at which a resource can be found
  - HTTP URL Syntax:

    `http://host_name/path/command?parameter1=value&parameter2...`

  - Example of HTTP URL:



HyperText Transfer Protocol (HTTP)

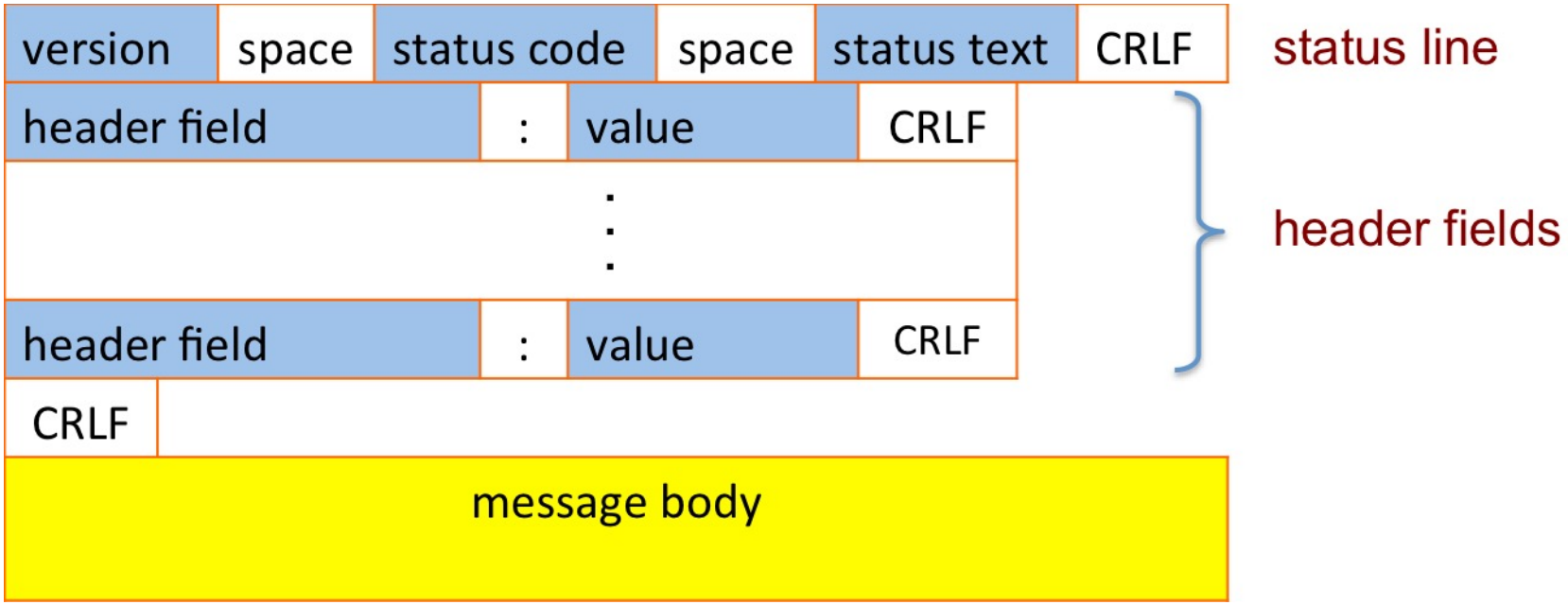
- Header fields: case insensitive

```
field name : field value
```

  - Host: host name from URL (required)
  - User-Agent: type of browser sending request
  - Accept: MIME types of acceptable documents (e.g text/html, image/jpg)
  - Content-Length: bytes in body
  - Content-Encoding: codings have been applied to the entity-body
  - Accept-Language: languages that are understandable by browser
- MINE: Convention for specifying content type of a message

HyperText Transfer Protocol (HTTP)

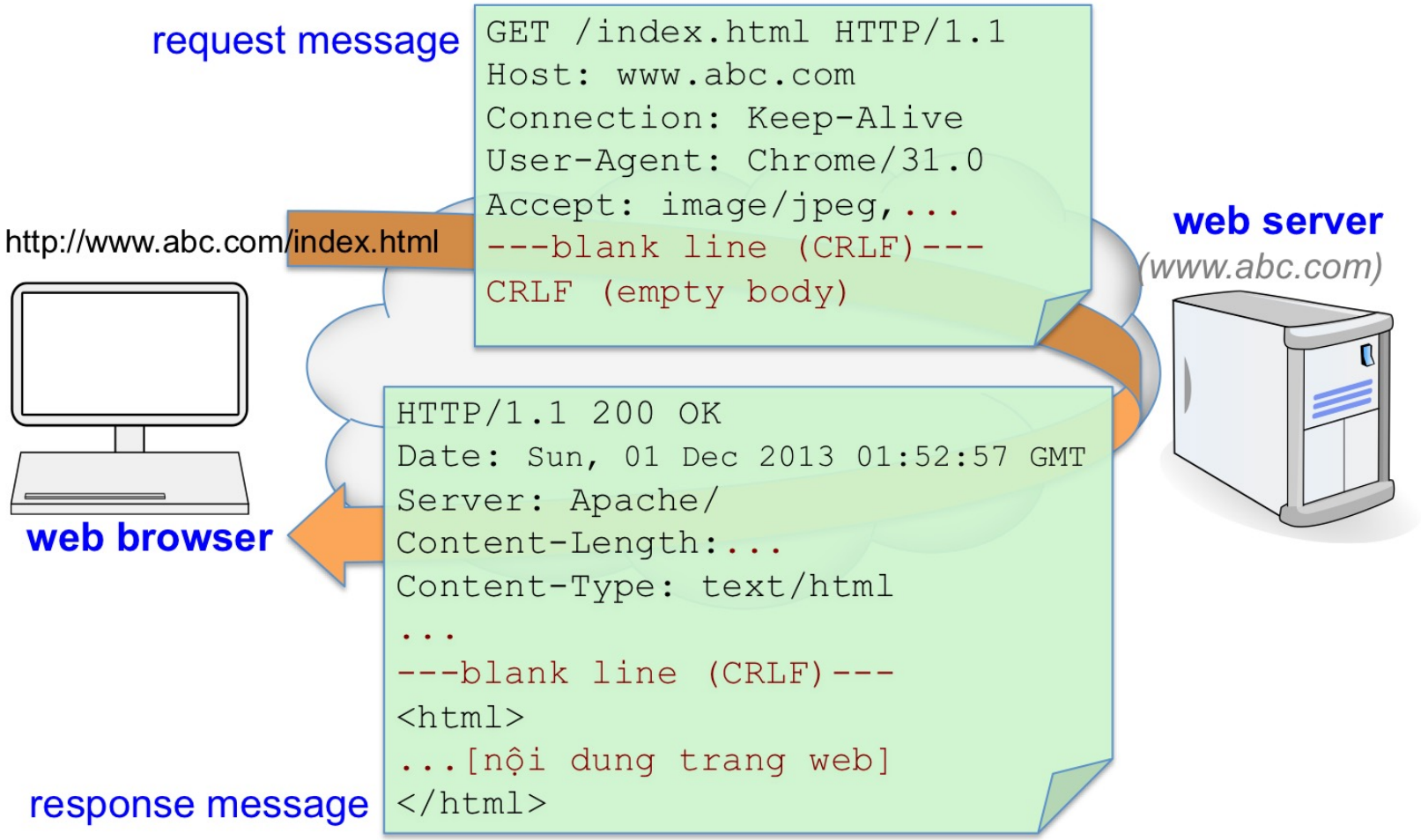
Structure of a HTTP response package:



HyperText Transfer Protocol (HTTP)

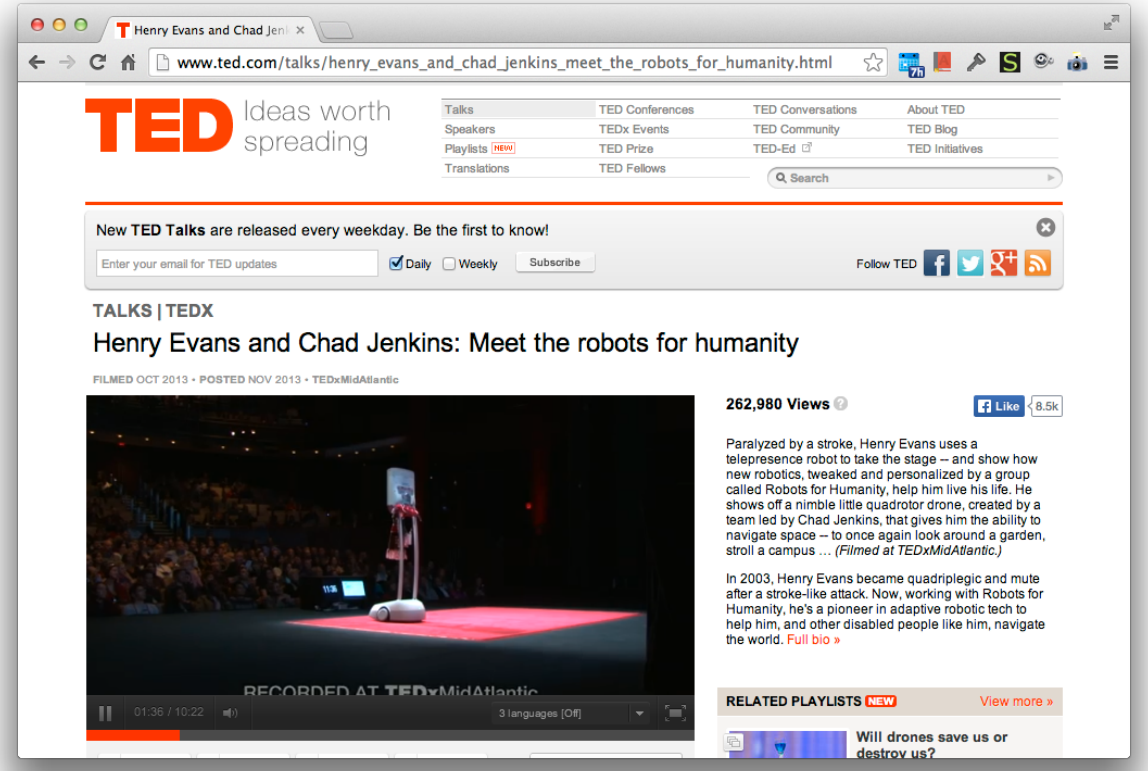
- Status: contains the status of the response
  - HTTP version
  - Status code: three digits
    - First digit: 1=information; 2=success; 3=redirection; 4=client error; 5=server error
    - Other two digits: additional information
  - Status text: reason phrase (for human use)
- Header fields:
  - Date: 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

Example of Request/Response method



WWW Resources

- Web pages
- Multimedia data: images, sound, videos, etc.



Uniform Resource Locator (URL)

- Used to identify a resource on the web
- Syntax:

```
<protocol>://<host address>[:port]/<path to the resource>
```

- Examples:
  - http://www.w3.org/Consortium/siteindex.html
  - ftp://ctan.tug.org/pub/tex/tds.zip
  - mailto:tcan@cit.ctu.edu.vn
  - file:///c:/Windows/clock.avi

Uniform Resource Locator (URL)

- Absolute URL:
  - Fully qualified URL which contains all elements of an URL
  - Example: http://www.w3.org/Consortium/siteindex.html
- Relative URL:
  - Contains only the relative path to the resource
  - Protocol and host address are inferred using the information in the URL of the containing page
  - Absolute path of the resource is calculated from the address (path) of containing webpage
  - Example: images/logo.png

Types of web

- Static web:
  - Content of the webpage is fixed
  - Extension: .html
  - Can be served directly by web server

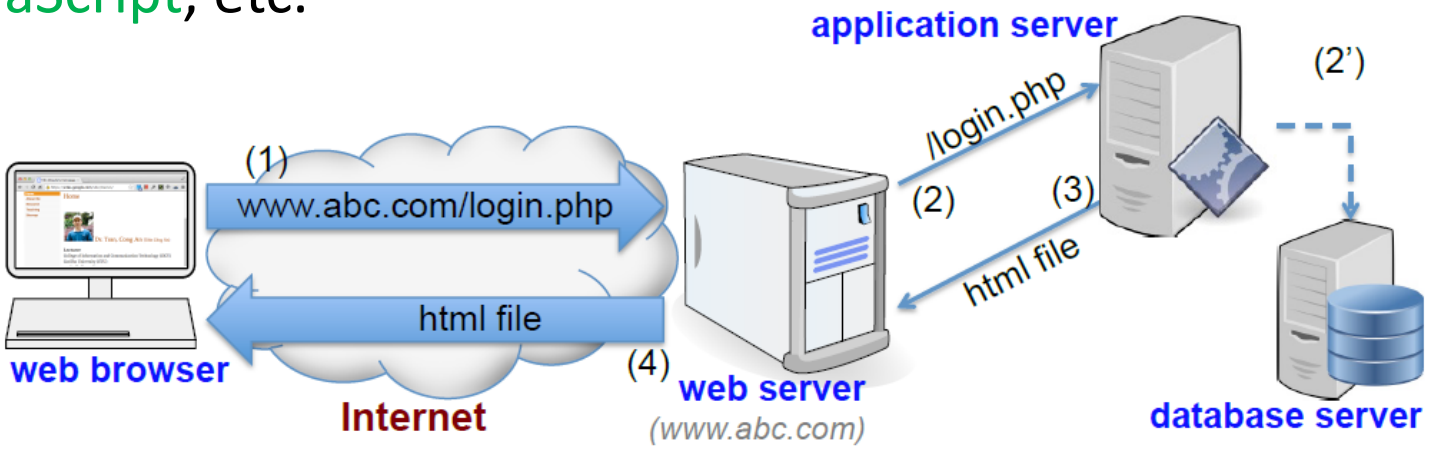
- Dynamic web:
  - Content of a webpage is “dynamic”
  - Usually produced by a programming/scripting language

Client-side scripting languages

- Embedded in the webpage
- Executed by browser
- Can be used to:
  - Interact with the web users
  - Validate data at the browser
  - Modify the web content or interface accordingly the context
  - Communicate with server
- Some client-side scripting languages: JavaScript, ActionScript, VBScript, etc.

Server-side scripting languages

- Used on the server side
- Do the business logic
- The result is usually a webpage
- Execute by an application server
- Some server-side scripting languages: ASP(.NET), JSP, PHP, JavaScript, etc.



HTML

(HyperText Markup Language)

Introduction to HTML

- HTML is the standard markup language for creating Web pages
- HTML describes the content and structure of Web pages using markup
- This is not a programming language!

- **1991**: HTML
- **1994**: HTML 2
- **1996**: CSS1 + JavaScript
- **1997**: HTML4 (*)
- **1998**: CSS2
- **2000**: XHTML 1
- **2002**: Tableless Web Design
- **2005**: AJAX
- **2009**: HTML 5

HTML Basic

- Structure of an HTML document:

```
<!DOCTYPE html>
<html>
    <head>
        head elements
    </head>

    <body>
        body elements
    </body>
</html>
```

Saved with **.html** extension

- An HTML document is made up of building blocks called elements

HTML Elements

- Elements are defined by tags (markers)
  - Opening tag: `<name>`
  - Closing tag: `</name>`
- The opening tag and its closing tag together specify a container for the content they enclose
- Not all tags have content: if a tag has no content, its form is `<name/>`
- The container and its content together are called an element
- Elements can be nested (element contains elements)

```
<p>This is a paragraph</p>
<img src="logo.png"/>
```

HTML Attributes

- An HTML element can have attributes
- Attributes provide additional information about an element
- Attributes are always specified in the start tag
- Attributes usually come in name/value pairs like:

```
name="value"
```

- Example:

```
<img src="img_girl.jpg" width="500" height="600">
```

Convention: use lowercase for attributes

HTML Document Head

- Contains metadata of the document: document title, character set, styles, links, scripts, and other meta information
- Defined by the `<head>` tag
- Placed between the `<html>` and `<body>` tags
- Typical tags used to describe the metadata: `<title>`, `<style>`, `<meta>`, `<script>`, and `<base>`

HTML Document Body

- Define the content of the document
- Basic content element
  - Headings: `<h1>, <h2>, …, <h6>`
  - Paragraphs: `<p>`
  - Quotations: `<blockquote>`
  - Links: `<a href="...">`
  - Images: `<img src="…">`
  - Tables: `<table>, <tr>, <th>, <td>`
  - Lists: `<ol>, <ul>, <li>, <dl>, <dt>, <dd>`
  - Blocks: `<div>, <span>`

HTML Formatting Elements

- `<b>`: **Bold text**
- `<strong>`: **Important text**
- `<i>`: *Italic text*
- `<em>`: *Emphasized text*
- `<small>`: Small text
- `<del>`: ~~Deleted text~~
- `<ins>`: Inserted text
- `<sub>`: Subscript text
- `<sup>`: Superscript text
- `<pre>`: pre-formatted text
- `<tt>`: typewriter-style text
- `<code>`: code style (monospace font)
- `<address>`: address section (usually italic)
- `<font color="…" face="…" size="…">`: color and font setting

HTML Styles

- Used to specify the style of an HTML element
- Can be done with the `style` attribute
- Syntax: `<tagname style="property:value;">`
  - `background-color`: specifies background color
  - `color`: specifies text colors
  - `font-family`: specifies text fonts
  - `font-size`: specifies text sizes
  - `text-align`: specifies text alignment

HTML Colors

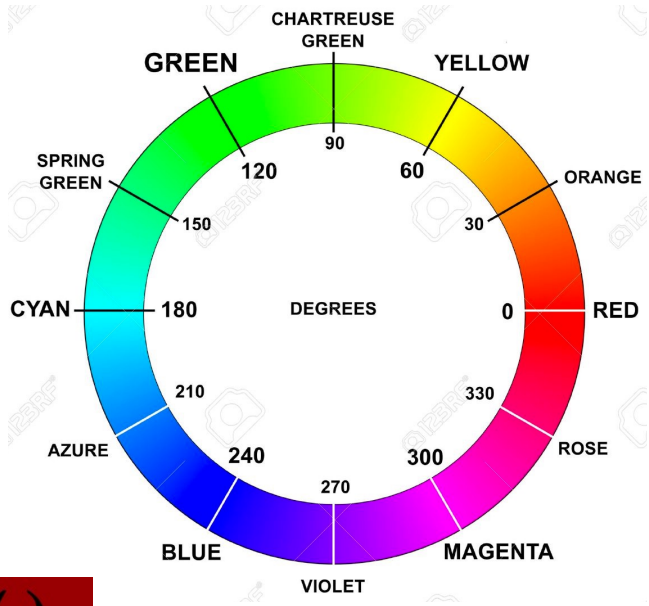
- Predefine color names: RED, GREEN, BLUE, BLACK, WHITE, etc. *(HTML supports 140 color names)*
- HEX: `#ffff00` (yellow)
- RGB: `rgb(255, 255, 0)`
- RGBA (RGB with opacity):

  rgba(255, 0, 0, 0.3) 0.5)

- HSL (hue, saturation, light):

  hsl(0, 100%, 50%) 80%) 30%)

- HSLA (hue, saturation, light, alpha)



HTML Forms

- Used to collect data from the web users
- Form data is posted to web server for processing (by an application server such as CGI, ASP, PHP, etc.)
- Tag: `<form action="server-side-script" method="POST/GET">`

HTML Forms Elements

- `<input type="…">`: used to accept data from users
  - `button`: a push button with no default behavior
  - `checkbox`: multiple values selection
  - `file`: file selection (to upload image to server)
  - `image`: graphical submit button
  - `password`: password input (value is obscured)
  - `radio`: single value selection
  - `reset`: resets the content of the form
  - `submit`: submit the form
  - `text`: single line textfield

HTML5: color, date, email, number, range, search, tel, time, url, week

HTML Forms Elements

- `<button type="…" onclick="…">`: clickable button with predefine standard functions
  - `type: submit, reset, button`
- `<select>`: provide a menu of options
- `<option>`: define items in a select element
- `<textarea rows="…", cols="…">`: multi-line plantext editor

Questions

Appendix

References

- http://www.w3.org(official)
  (http: // www. w3. org/ community/ webed/ wiki/ Main_ Page)
- http://w3school.com/html/(recommended)
- http://htmldog.com/guides/html/
- http://www.echoecho.com/html.htm

Quiz

1. **What does HTML stand for?**
   a. Hyperlinks and Text Markup Language
   b. Home Tool Markup Language
   c. Hyper Text Markup Language

2. **Who is making the Web standards?**
   a. The World Wide Web Consortium
   b. Mozilla
   c. Microsoft
   d. Google

Quiz

3. **Choose the correct HTML tag for the largest heading**
   a. <h1>
   b. <heading>
   c. <head>
   d. <h6>

4. **Choose the correct HTML tag to make a text bold**
   a. <bold>
   b. <b>
   c. <i>
   d. <u>

Quiz

5. **What is the correct HTML for creating a hyperlink?**
   a. <a>http://www.w3schools.com<a>
   b. W3Schools.com
   c. W3Schools.com
   d. W3Schools

6. **How can you create an e-mail link?**
   a. <mail>xxx@yyy</mail>
   b. <a href="mailto:xxx@yyy">
   c. <mail href="xxx@yyy">
   d. <a href="xxx@yyy">

Quiz

7. **How can you open a link in a new browser window?**
   a. <a href="url" target="_blank">
   b. <a href="url" target="new">
   c. <a href="url" new>

8. **Which of these tags are all table tags?**
   a. <thead>, <body>, <tr>
   b. <table>, <tr>, <tt>
   c. <table>, <head>, <tfoot>
   d. <table>, <tr>, <td>

Quiz

9. **What is the correct HTML for inserting an image?**
   a.
   b.
   c. <image src="image.gif" alt="MyImage"/>
   d. <img alt="MyImage">image.gif</img>

10. **Why should you add alternative text to your images?**
   a. So the user can save the image using the text as a name
   b. So the users can get an idea of what the image is before it loads
   c. In case the user wishes to load a different picture
   d. So the users can get an idea of what the image is in case the browser fails to load the image

Quiz

11. **To seperate single list items use?**
    a.
    b.
    c.

12. **When making bulleted lists you have what options?**
    a. triangle, square, circle
    b. square, disc, polygon
    c. disc, circle, square

Quiz

13. **What are the fields that allows the visitor to enter information called?**
    a. Meta tags
    b. Form fields
    c. Meta fields

14. **The `value` setting of a text field defines what?**
    a. The length of the field
    b. If the value entered is a proper value
    c. What will appear in the field as the default value

Quiz

15. **Which field can hold information that does not show?**
    a. Text field
    b. Hidden field
    c. Frame field

16. **Password fields are similar to what?**
    a. Hidden fields except text show as "*"
    b. Text fields except text show as "*"
    c. Text area except text show as "*"

HTML5

- The latest version of the HTML markup language (5.0, 5.1, 5.2)
- The living standard
- The next generation features for modern web development
- Theoretically, the HTML5 specification will be completed in 2022.

HTML5 $\sim=$ HTML + CSS + JS

HTML5 Characteristics

- Simpler and cleaner
  minimize syntax for tags and attributes
- New structural and semantic elements
  richer, more meaningful and more flexible
- New form elements and attributes
  13 new input types and many new form elements
- New other elements
  multimedia content (video, audio, canvas, menu, etc.)
- New functions
  web storage, real-time communication, drag and drop, geolocation, etc.