



CT428. LẬP TRÌNH WEB

CHƯƠNG 1. GIỚI THIỆU DỊCH VỤ WWW

Giảng viên: Trần Công Ân (tcan@cit.ctu.edu.vn)
<https://sites.google.com/site/tcanvn/>

Bộ môn Mạng máy tính và Truyền thông
Khoa Công Nghệ Thông Tin & Truyền Thông
Đại học Cần Thơ

2014

NỘI DUNG

GIỚI THIỆU INTERNET VÀ DỊCH VỤ WWW

CÁC THÀNH PHẦN CỦA DỊCH VỤ WWW

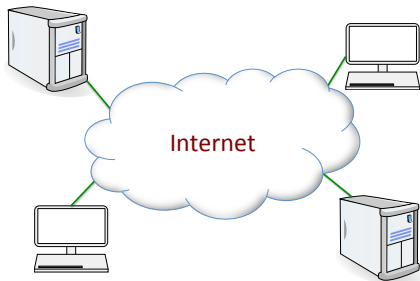
GIAO THỨC HTTP

ĐỊNH VỊ TÀI NGUYÊN TRÊN WWW

CÁC LOẠI SIÊU VĂN BẢN

INTERNET LÀ GÌ?

- ▶ là một **mạng toàn cầu**, nối kết các mạng máy tính sử dụng bộ giao thức (protocol) TCP/IP
- ▶ tiền thân là mạng **ARPANET** của Bộ quốc phòng Mỹ (60s)
- ▶ mục tiêu ban đầu là dùng để **trao đổi tập tin** (FTP) và **thư điện tử** (email)
- ▶ được thương mại hoá vào cuối những năm 80s
- ▶ một số **dịch vụ cơ bản khác** trên internet: telnet, usenet, chat, WWW, ...



CÁC ĐẶC ĐIỂM CƠ BẢN CỦA MẠNG INTERNET

- ▶ [hầu như] không có sự kiểm soát tập trung (**không có ai sở hữu** toàn bộ mạng Internet)
⇒ các “mạng con” có thể hoạt động độc lập (tương đối)
- ▶ được xây dựng dựa trên các **chuẩn “mở”**
⇒ tất cả mọi người đều có thể tạo ra các thiết bị hay dịch vụ cho Internet
- ▶ các máy tính hoặc mạng máy tính có thể **gia nhập hoặc rời** mạng Internet một cách “tự do”
- ▶ các dịch vụ Internet được tổ chức theo **mô hình client/server**

CÁC THUẬT NGỮ/VIẾT TẮT THƯỜNG GẶP

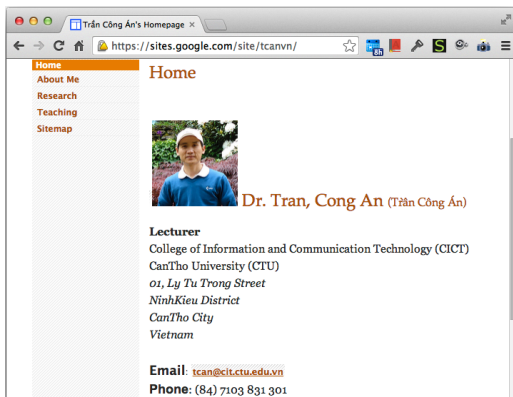
- ▶ **TCP/IP** (Transmission Control Protocol/Internet Protocol): bộ giao thức Internet.
- ▶ **DNS** (Domain Name System): hệ thống tên miền.
- ▶ **Các giao thức**: HTTP (WWW), SMTP/POP/IMAP (email), FTP (file transfer), ...
- ▶ **URL** (Uniform Resource Locator): bộ định vị tài nguyên đồng dạng.
- ▶ **Server**: chương trình cung cấp dịch vụ Internet (*hay máy tính cài đặt các dịch vụ*).
- ▶ **Client**: chương trình sử dụng dịch vụ Internet (*hay máy tính sử dụng để truy cập các dịch vụ Internet*).

CÁC TỔ CHỨC QUAN TRỌNG CỦA INTERNET

- ▶ Internet Engineering Task Force (IETF): quản lý **giao thức** chuẩn của Internet.
- ▶ Internet Corporation for Assigned Names and Numbers (ICANN): quản lý **hệ thống tên miền** cấp cao nhất.
- ▶ World Wide Web Consortium (W3C): quản lý chuẩn của **dịch vụ WWW**.

WORLD WIDE WEB (WWW) LÀ GÌ?

- ▶ là một **dịch vụ** trên Internet, dùng để **trao đổi** hay chia sẻ thông tin dưới dạng **siêu văn bản** (hypertext)
- ▶ được tạo ra bởi Tim Berners-Lee (CERN) vào đầu những năm 90s (1989-1991)
- ▶ là hình thức trao đổi/chia sẻ thông tin tiện lợi và phổ biến trên Internet



SIÊU VĂN BẢN, TRANG WEB, WEBSITE

- ▶ **Siêu văn bản:** là các văn bản có chứa các **siêu liên kết** (hyperlink), cho phép:
 - ▶ **liên kết** đến các siêu văn bản hay các nguồn tài nguyên khác (văn bản, hình ảnh, âm thanh, ...) trên Internet
 - ▶ **truy xuất đến các siêu văn bản** khác theo cách không tuần tự
 - ▶ [thông thường] có phần mở rộng là **.html** hoặc **.htm**
- ▶ **Trang web (web page):** là một siêu văn bản.
- ▶ **Website:** là một **hệ thống các trang web có liên quan với nhau** (của một tổ chức hay cá nhân). VD: website Trường ĐHCT, website Khoa CNTT&TT, ...

CÁC THÀNH PHẦN CỦA DỊCH VỤ WWW

WWW, cũng như tất cả các dịch vụ khác trên Internet, được tổ chức theo mô hình **client/server** (khách hàng/phục vụ)

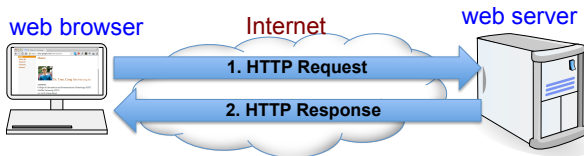
- ▶ **Server**: là một chương trình **cung cấp dịch vụ** trên Internet
 - ▶ “lắng nghe” yêu cầu từ các clients
 - ▶ xử lý và phản hồi các yêu cầu của các clients
 - ▶ server trong dịch vụ WWW được gọi là **web server**
- ▶ **Client**: là một chương trình **sử dụng dịch vụ** trên Internet
 - ▶ gửi yêu cầu đến server
 - ▶ nhận và xử lý kết quả trả về (hiển thị cho người dùng, ...)
 - ▶ một client trong dịch vụ WWW được gọi là **web browser**

WEB SERVER

- ▶ là một chương trình **phục vụ yêu cầu về web** (yêu cầu trang web)
 - ▶ lưu trữ các trang web
 - ▶ nhận yêu cầu về web từ web client (web browser)
 - ▶ xử lý và gửi kết quả (trang web) về cho web client
- ▶ **cổng (port) mặc định** của các web server là 80
- ▶ một số web server thông dụng: Apache, Microsoft Internet Information Service (IIS)

WEB BROWSER (TRÌNH DUYỆT WEB)

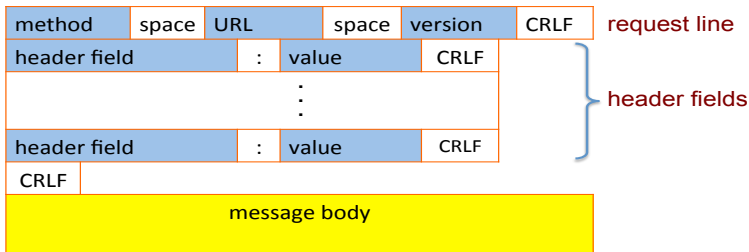
- ▶ là chương trình **sử dụng dịch vụ WWW**:
 - ▶ gửi yêu cầu trang web (của người dùng) đến web server
 - ▶ nhận và hiển thị trang web trả về từ web server
- ▶ một số trình duyệt web thông dụng: Internet Explore, Chrome, Firefox, Safari, Opera, ...



GIAO THỨC HTTP

- ▶ viết tắt của **HyperText Transfer Protocol**, là **giao thức của dịch vụ WWW**.
- ▶ qui định **cách thức “giao tiếp”** giữa web server và web browser
- ▶ bao gồm một **tập các lệnh và “qui ước”** chung giữa web server và web browser dùng để trao đổi dữ liệu
- ▶ dữ liệu trao đổi chủ yếu là các **siêu văn bản**
- ▶ trải qua nhiều phiên bản: HTTP 0.9, HTTP 1.0, HTTP 1.1

CẤU TRÚC THÔNG điệp YÊU CẦU



- ▶ method: phương thức yêu cầu
- ▶ URL: đường dẫn đến đối tượng liên quan đến yêu cầu
- ▶ version: HTTP version mà client muốn giao tiếp với server (1.0, 1.1)
- ▶ header fields: mô tả các thông tin khác liên quan đến yêu cầu của client

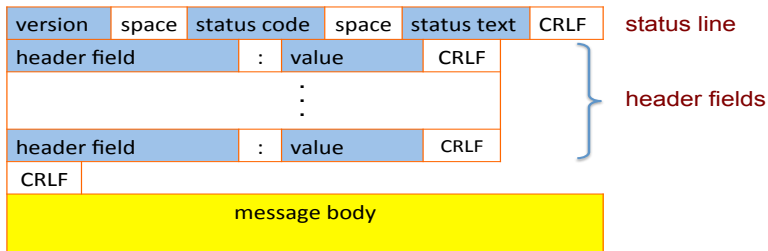
CÁC PHƯƠNG THỨC YÊU CẦU CƠ BẢN

- ▶ **GET**: **yêu cầu một trang web** từ server (chỉ định bằng URL). Đối số cung cấp cho web server (nếu có) sẽ được **truyền thông qua URL**.
- ▶ **POST**: yêu cầu một trang web từ server (chỉ định bằng URL). Đối số cung cấp cho web server (nếu có) sẽ được truyền tách biệt với URL, **đặt bên trong thông điệp yêu cầu (message body)**.
- ▶ **PUT**: **upload** một trang web lên server, đặt tại vị trí được chỉ định bởi URL.

CÁC HEADER FIELDS THÔNG DỤNG

- ▶ dùng để chỉ định một số **ràng buộc đối với web server** trong việc phục vụ yêu cầu của web browser
- ▶ **cú pháp** của mỗi field: <tên field>: <giá trị> [,<giá trị>]*
- ▶ một số fields cơ bản (phân biệt chữ hoa, chữ thường):
 - ▶ Accept: loại media được chấp nhận bởi browser (text/html, image/jpeg)
 - ▶ Content-Length: kích thước của message body (bytes)
 - ▶ Content-Encoding: mã hóa của message body (gzip, ...)
 - ▶ Accept-Language: ngôn ngữ chấp nhận bởi browser (vn, en)

CẤU TRÚC THÔNG điệp ĐÁP ỨNG

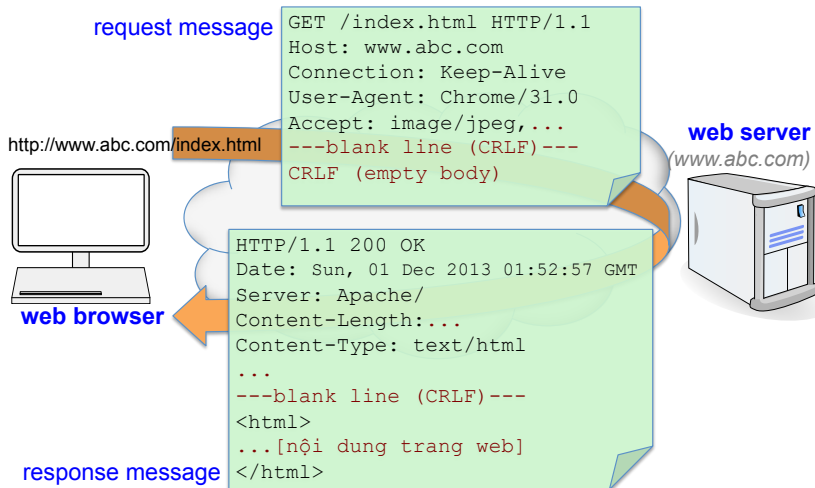


- ▶ **status:** bao gồm 3 chữ số, cho biết **trạng thái phục vụ** yêu cầu của web browser
- ▶ **header fields:** **chứa thông tin về web server** và các thông tin liên quan đến trang web được yêu cầu (tương tự thông điệp yêu cầu)

MỘT SỐ MÃ TRẠNG THÁI CƠ BẢN

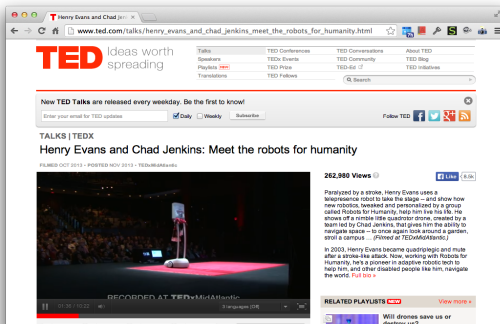
- ▶ **1xx:** **Thông tin** – Yêu cầu đã được nhận, đang tiếp tục xử lý
- ▶ **2xx:** **Thành công** – Yêu cầu đã được nhận thành công, có thể xử lý hay đã được chấp nhận
- ▶ **3xx:** **Chuyển hướng** – Cần thực hiện thêm một số thao tác để server có thể hoàn tất phục vụ yêu cầu
- ▶ **4xx:** **Lỗi client** – Yêu cầu không hợp lệ hoặc không thể được đáp ứng
- ▶ **5xx:** **Lỗi server** – Server không thể đáp ứng yêu cầu

MINH HỌA THÔNG điệp YÊU CẦU/ĐÁP ỨNG



TÀI NGUYÊN TRÊN WWW

- ▶ Tài nguyên trong dịch vụ WWW:
 - ▶ các trang web
 - ▶ dữ liệu đa phương tiện (multimedia) như: âm thanh, hình ảnh, ...



BỘ ĐỊNH VỊ TÀI NGUYÊN ĐỒNG DẠNG (URL)

- ▶ Bộ định vị tài nguyên đồng dạng (**U**niform **R**esource **L**ocator - URL) được dùng để **định vị các tài nguyên** trên WWW (và Internet).
- ▶ Một URL cơ bản của dịch vụ WWW có dạng như sau:

`http://www.w3.org/Consortium/siteindex.html`

giao thức địa chỉ server đường dẫn

- ▶ **Tổng quát:**

`<giao thức>://<địa chỉ host>[:port]/<đường dẫn>`

CÚ PHÁP URL CỦA MỘT SỐ DỊCH VỤ CƠ BẢN

- ▶ **HTTP:** `http://<địa chỉ host>[:port]/<path>`
- ▶ **FTP:**
`ftp://[uname[:pass]@]<địa chỉ host>[:port][/path]`
 - ▶ `ftp://ctan.tug.org/`
 - ▶ `ftp://ctan.tug.org/pub/tex/tds.zip`
- ▶ **Email:** `mailto:<email address>`
 - ▶ `mailto:tcn@cit.ctu.edu.vn`
- ▶ **File:** `file://<đường dẫn đến file cần truy xuất>`
 - ▶ `file://192.186.1.106/Users/Shared/readme.txt`
 - ▶ `file:///c:/Windows/clock.avi`

URL TƯƠNG ĐỐI

- ▶ là địa chỉ **không đầy đủ** tới một tài nguyên
- ▶ chỉ bao gồm **đường dẫn** đến tài nguyên cần truy xuất, không có giao thức và địa chỉ host
- ▶ giao thức và địa chỉ host được **suy ra** từ trang web chứa URL tương đối
- ▶ ví dụ, trong một trang web có URL
"http://www.w3c.org/Consortium/siteindex.html":
 - ▶ URL "membership.html" sẽ **tương đương** với:
"http://www.w3c.org/Consortium/membership.html"
 - ▶ URL "/Help/Webmaster.html" sẽ **tương đương** với:
"http://www.w3c.org/Help/Webmaster.html"

CÁC LOẠI SIÊU VĂN BẢN

- ▶ Web tĩnh (static web):
 - ▶ là những trang web có **nội dung cố định**, không thay đổi theo ngữ cảnh
 - ▶ có phần mở rộng là **.html** hoặc **.htm**
- ▶ Web động (dynamic web):
 - ▶ là những trang web có **nội dung thay đổi tùy theo ngữ cảnh** (đối số kèm theo yêu cầu của người dùng)
 - ▶ nội dung được tạo ra bởi một **ngôn ngữ script**, bao gồm 2 loại:
 - ▶ **client-side**: mã script được thực thi bởi web browser
 - ▶ **server-side**: mã script được thực thi bởi web server (hay một application server)

NGÔN NGỮ SCRIPT CLIENT-SIDE

- ▶ được “nhúng” vào trong **tập tin html**
- ▶ được **thực thi bởi web browser** sau khi nhận trang web về từ web server
- ▶ đoạn script này **có thể thay đổi nội dung của trang web** tùy vào ngữ cảnh (context)
- ▶ một số ngôn ngữ script client-side: JavaScript, ActionScript, VBScript, ...

NGÔN NGỮ SCRIPT SERVER-SIDE

- ▶ là đoạn script dùng để tạo ra một trang web
- ▶ được thực thi bởi web server (thường là bởi 1 application server)
- ▶ sau khi script được thực thi, trang web tạo ra bởi script sẽ được web server gửi trả về cho web browser
- ▶ một số ngôn ngữ script server-side: ASP(.NET), JSP, PHP, ...

