中山大學
SUN YAT-SEN UNIVERSITY

Lecture 1
Internet and World Wide Web

SE-805 Web 2.0 Programming (supported by Google)

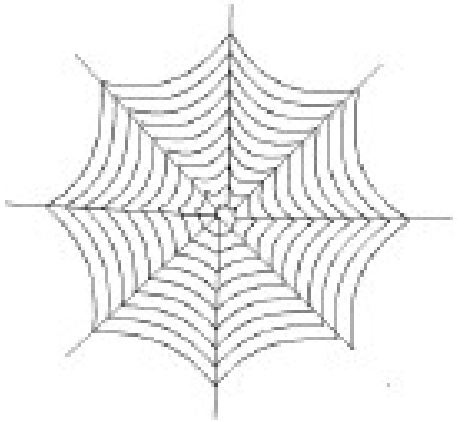
http://my.ss.sysu.edu.cn/courses/web2.0/

School of Software, Sun Yat-sen University

Outline

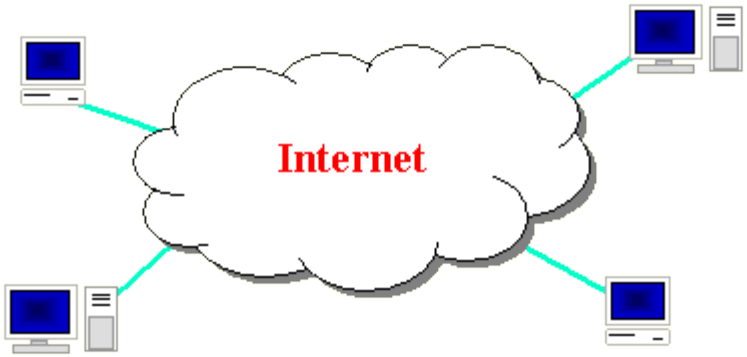
- **The Internet**
- The World Wide Web (WWW)
- Web 2.0

What's the Internet?



- A Chinese town officer
  - *"The Internet is an English agents' net"*
- A U.S. Senator
  - *"series of tubes" (explanation)*
- How many internets are there, anyway? Is The Google one of them?

The Internet



- Wikipedia: http://en.wikipedia.org/wiki/Internet
- A connection of computer networks using the Internet Protocol Suite (TCP/IP)
- What's the difference between the Internet and the World Wide Web (WWW)?
- WWW = HTML* + HTTP(S)

* *Including CSS, JavaScript, and other browser enabled contents*

Brief History

- Began as a US Department of Defense network called ARPANET (1960s-70s)
- Initial services: electronic mail, file transfer
- Opened to commercial interests in late 80s
- WWW created in 1989-91 by Tim Berners-Lee
- Popular web browsers released: Netscape 1994, IE 1995
- Amazon.com opens in 1995; Google in1996
- Chinese First Connection with Internet：Chinese Academics Net, by Computer Applying Technology Institute of Beijing1986
- First email, Sep. 14 1987, from CATIB, "Across the Great Wall we can reach every corner in the world"
- Chinese First Full Internet Connection：NCFC (National Computing and Networking Facility of China) 1994
- Baidu in1999；Taobao in 2003

Key Aspects of the Internet

- Internet is for information sharing
- “internet” vs. “Internet”
- Subnetworks can stand on their own
- Computers can dynamically join and leave the network
- Built on open standards; anyone can create a new device
- Lack of centralized control (mostly)
- Everyone can use it with simple, commonly available software

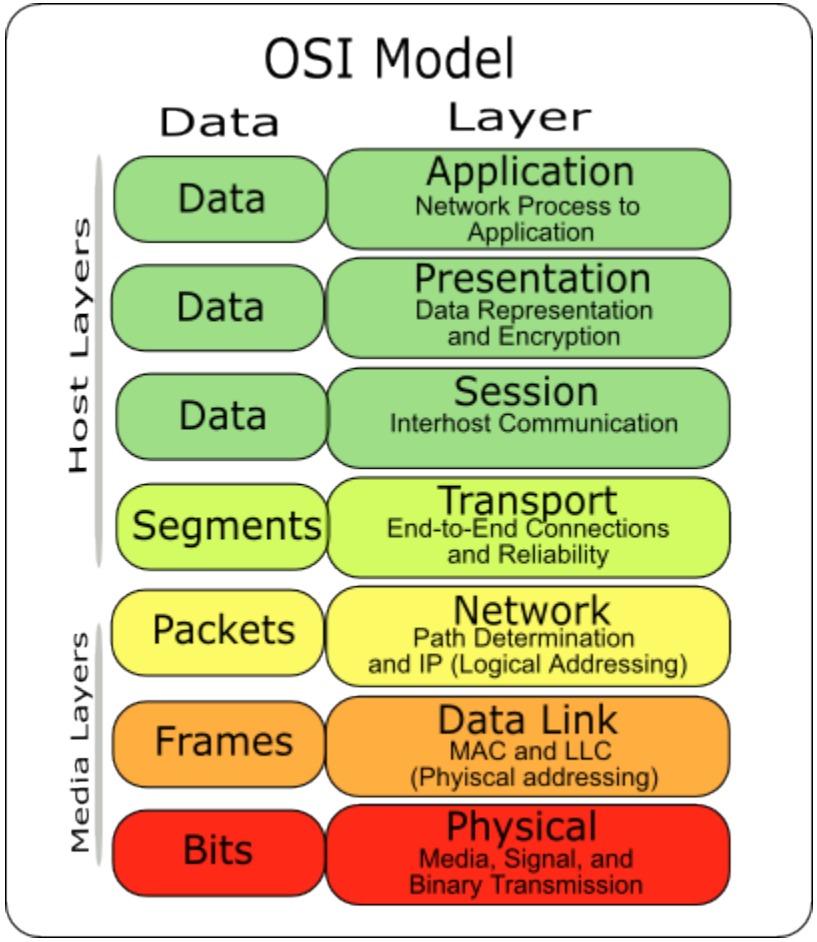
People and Organizations

- Internet Engineering Task Force (IETF): internet protocol standards
- Internet Corporation for Assigned Names and Numbers (ICANN): decides top-level domain names
- World Wide Web Consortium (W3C): Web standards

Layered Architecture

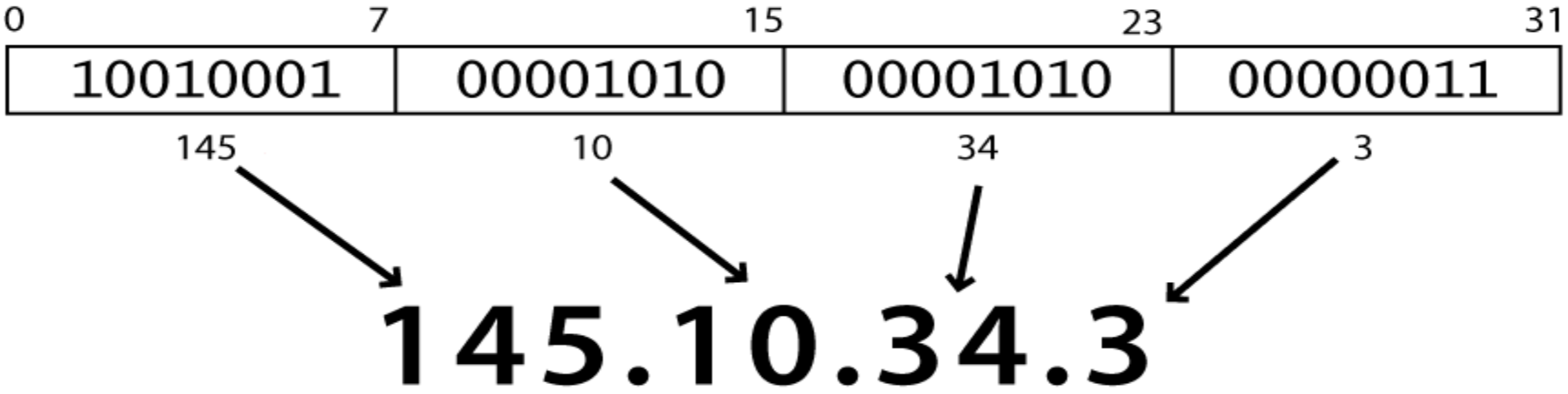
The internet uses a layered hardware/software architecture (OSI model):

- **Physical layer**: devices such as coaxial cables, fiber-optic lines, modems
- **Data link layer**: basic hardware protocols (Ethernet, Wi-Fi, DSL, ATM, PPP)
- **Network / internet layer**: basic software protocols (i.e. IP)
- **Transport layer**: adds reliability to network layer (TCP, UDP)
- **Application layer**: implements specific communication for each kind of program (HTTP, POP3/IMAP, SSH, FTP)



Internet Protocol (IP)

- **IP** is the underlying protocol of communication for all data (packets) sent across the internet.
- Each device has a 32-bit IP address, shown as four 8-bit numbers (0-255)



- Find out your internet IP address: whatismyip.com
- Find out your local IP address:
  - in a terminal, type: ipconfig (Windows) or ifconfig (Mac/Linux)
- IP v4 vs. IP v6 (32-b vs. 128-b)

Transmission Control Protocol (TCP)

- Adds multiple, guaranteed message delivery on top of IP
- **Multiplexing**: multiple programs using the same IP address
  - **Port**: a number given to each program or service
    - 80: Web browser (443 for secure browsing)
    - 25: Email
    - 22: SSH
    - 21: FTP
    - More common ports
- Some programs (QQ, games, streaming media programs) use simpler UDP protocol instead of TCP
- Find out ports used:
  - In a terminal, using **netstat** (Windows) command
  - Using CurrPorts

Outline

- The Internet
- **The Word Wide Web (WWW)**
- Web 2.0

Web Servers and Browsers

- **Web server**: software that listens for Web page requests
  - Apache
  - Microsoft Internet Information Server (IIS) (part of Windows)
- **Web browser**: fetches/displays documents from Web servers
  - Microsoft Internet Explorer (IE)
  - Mozilla Firefox
  - Apple Safari
  - Google Chrome
  - Opera

Domain Name System (DNS)

- A set of servers that map written names to IP addresses
  - Example: www.sysu.edu.cn → 202.116.64.9
  - Using Windows command **nslookup** to find out IP address
  - Non-english languages in domain name IDN ccTLD Fast Track
- Many systems maintain a local cache called a host file
  - Windows: C:\Windows\system32\drivers\etc\hosts
  - Mac: /private/etc/hosts
  - Linux: /etc/hosts

Uniform Resource Locator (URL)

- An identifier for the location of a document on a web site
- A basic URL:

  http://www.aw-bc.com:80/info/regesstepp/index.html

  ~~~ ~~~~~~~~~~~~~ ~~ ~~~~~~~~~~~~~~~~~~~~

  protocol host port path

- Upon entering this URL into the browser, it would:
  - Ask the DNS server for the IP address of `www.aw-bc.com`
  - Connect to that IP address at port 80
  - Ask the server to GET `/info/regesstepp/index.html`
  - Display the resulting page on the screen
  ~~~

More Advanced URLs

- **Anchor**: jumps to a given section of a web page
  http://www.textpad.com/download/index.html**#downloads**
  - Fetches index.html then jumps down to the part of the page labeled downloads

- **Port**: for web servers on ports other than the default 80
  http://www.cs.washington.edu**:8080**/secret/money.txt

- **Query string**: a set of parameters passed to a web program
  http://www.google.com/search**?q=miserable+failure&start=10**
  - Parameter q is set to "miserable+failure"
  - parameter start is set to 10

HyperText Transport Protocol (HTTP)

- The set of commands understood by a web server and sent from a browser
- Some HTTP commands (your browser sends these internally):
  - GET *filename* : download
  - POST *filename* : send a web form resquest
  - PUT *filename* : upload
  - DELETE *filename*: remove entity
  - HEAD *filename*: only status information, not entire content
- Simulating a browser with a terminal window:

```
$ telnet www.sysu.edu.cn 80
Trying 202.116.64.9... Connected to 202.116.64.9
(202.116.64.9). Escape character is '^]'.
GET /2009/xxgk.html
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0
..."> <html> ...
```

HTTP Error Codes

- When something goes wrong, the web server returns a special "error code" number to the browser, possibly followed by an HTML document
- Common error codes:

Number	Description
200	OK
301-303	page has moved (permanently or temporarily)
403	you are forbidden to access this page
404	page not found
500	internal server error
complete list	

Internet Media (**MIME**) Types

- Sometimes when including resources in a page (style sheet, icon, multimedia object), we specify their type of data

MIME Type	File Extension
text/html	.html , .htm, shtml, .shtm
text/plain	.txt
image/gif	.gif
image/jpeg	.jpg
video/quicktime	.mov
application/octet-stream	.exe

- Lists of MIME types: by type, by extension
- ".html" vs. ".htm"

Web Languages / Technologies

- HyperText Markup Language (HTML): used for writing web pages
- Cascading Style Sheets (CSS): stylistic info for web pages
- PHP HyperText Processor (PHP): dynamically create pages on a web server *– of course, there are many other languages and scripts that can do this …*
- JavaScript: interactive and programmable web pages
- Asynchronous JavaScript and XML (AJAX): accessing data for web applications
- eXtensible Markup Language (XML): meta-language for organizing data
- Structured Query Language (SQL): interaction with databases
- Resource Description Frame (RDF): describing web resources semantically
- ……

Terms

- **Internet Service Provider (ISP)**
  - Enterprises or organizations who provide Internet access to you,
  - Who? please identify your ISPs
- **Web Hosting**
  - Provide a place for consumers to store pages designed to be consumed by the Web surfing public
  - ISPs often offer Web hosting services along with their standard connectivity packages.
- **Client/Server** vs. **Browser/Server**
- **Presentation Layer**
  - Often refers to the top tier of enterprise application architecture
  - in Web, it includes both code of Web pages and code creating Web pages.
- **Client Side Scripting/Programming**
  - Writing code consumed by browsers to render Web pages and to respond to users' interactions
- **Server Side Scripting/Programming**
  - Writing code used to generate the code consumed by browsers

Outline

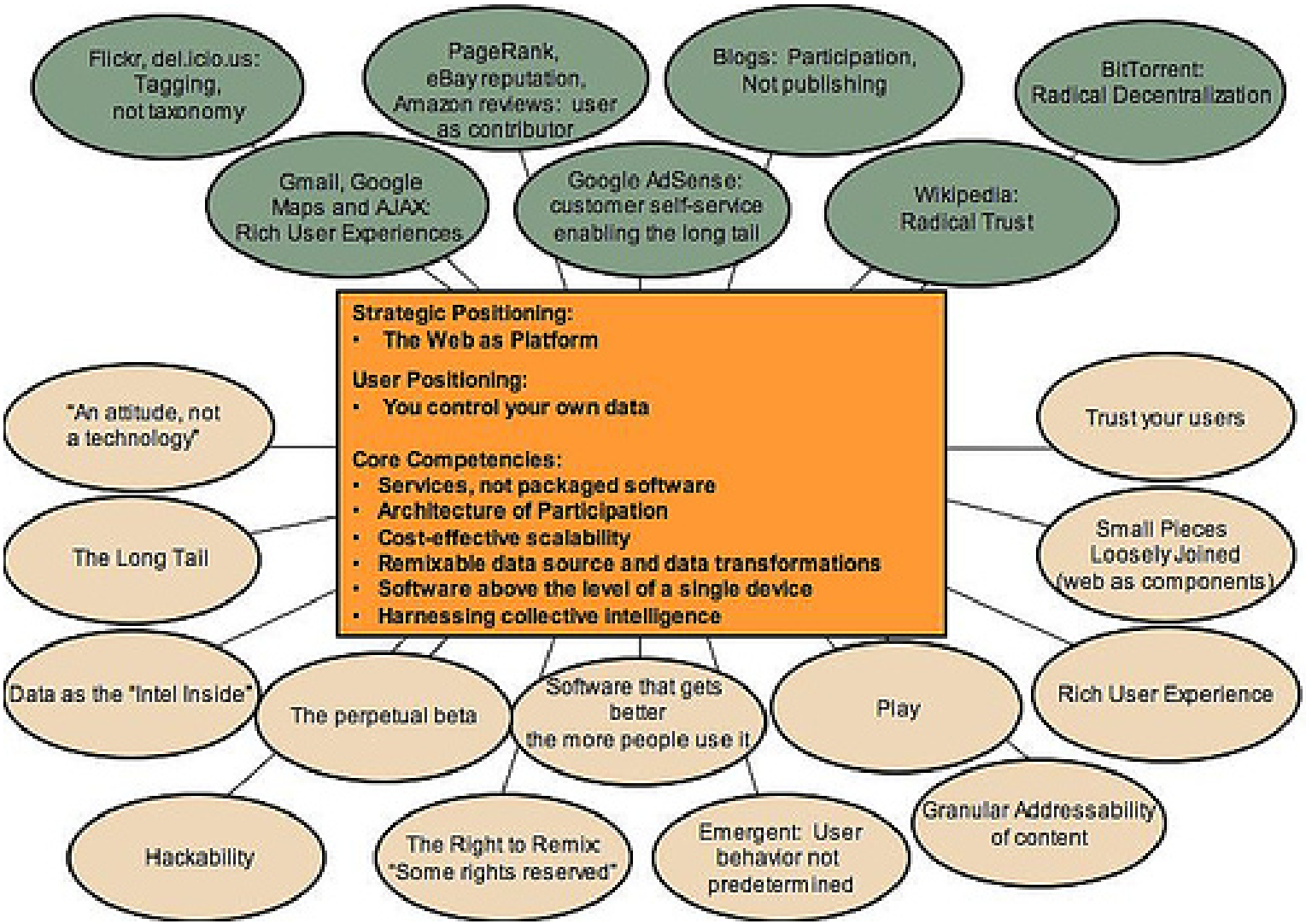
- The Internet
- The Word Wide Web (WWW)
- **Web 2.0**

Web 1.0 vs. Web 2.0

- **Web 1.0** is about publishing
  - Users are limited to the passive browsing of information that is provided to them
- **Web 2.0** is about interaction
  - Allows its users to interact with other users or to change website content
  - Information sharing, interoperability, user-centered design and collaboration
  - Hosted services, web applications, social-networking sites, video-sharing sites, wikis, blogs, mashups and folksonomies.
  - Coined by Tim O'Reilly in the O'Reilly Media Web 2.0 conference in 2004

Web 2.0 Memo Map

http://oreilly.com/web2/archive/what-is-web-20.html



Web 2.0 Examples

http://oreilly.com/web2/archive/what-is-web-20.html

Web 1.0		Web 2.0
DoubleClick	-->	Google AdSense
Ofoto	-->	Flickr
Akamai	-->	BitTorrent
mp3.com	-->	Napster
Britannica Online	-->	Wikipedia
personal websites	-->	blogging
evite	-->	upcoming.org and EVDB
domain name speculation	-->	search engine optimization
page views	-->	cost per click
screen scraping	-->	web services
publishing	-->	participation
content management systems	-->	wikis
directories (taxonomy)	-->	tagging ("folksonomy")
stickiness	-->	syndication

2.0 Flurry

- Library 2.0, Classroom 2.0, Publishing 2.0, Learning 2.0
- Social Work 2.0, Enterprise 2.0, PR 2.0
- Medicine 2.0, Telco 2.0, Travel 2.0
- Government 2.0
- and even Porn 2.0
- These 2.0s refer to Web 2.0 technologies as the source of the new version in their respective disciplines and areas.

Web 2.0 Technologies

- **Browser side**
  - Asynchronous JavaScript and XML (AJAX),
  - RIA
    - Adobe Flash
    - JavaScript / AJAX frameworks
      - Prototype, script.aculo.us, Yahoo! UI Library, Dojo Toolkit, MooTools, jQuery, ExtJS, …
    - Others
      - XUL, JavaFX, Silverlight, OpenLaszlo, …
- **Server side**
  - many of the same technologies as Web 1.0
    - PHP, Ruby, ColdFusion, Perl, Python, JSP, Servlet, and ASP
  - More different formats
    - XML, RSS, and JSON , why?

Summary

- The Internet
  - History
  - Key aspects
  - People and organizations
  - Layered architecture
  - Protocols: IP, TCP
- The World Wide Web (WWW)
  - Servers and browsers
  - Protocols: DNS, URL, HTTP, MIME
  - Web langauges / technologies
- Web 2.0
  - Features, advantages, applications, and technologies

Exercises

- Use a terminal shell on your computer to fetch the homepage of the School of Software, Sun Yat-sen University
- Install the Firefox and the Firebug add-on

Reading materials

- A Brief Histroy of the Internet
  http://www.isoc.org/internet/history/brief.shtml
- Introduction to Web 2.0
  http://en.wikipedia.org/wiki/Web_2.0 ,
  http://oreilly.com/web2/archive/what-is-web-20.html

Thank you!