# **ITEC 4020**

**Instructor: Jimmy Huang** 

### **ProfJimmyHuang@gmail.com**

https://yorku.zoom.us/j/94310213415?pwd=UC9MeVZvdTAvT29XM113aGJRVjAwdz09 Meeting ID: 943 1021 3415 Passcode: 956408

http://www.yorku.ca/jhuang/4020A.html

# Motivation

### Web-based Knowledge & Data Management

- A huge amount of Web data
- > how to organize, retrieve them, how to discover interesting patterns and how to make a recommend from them
- ⇒ Web Search Engine
- Uber Taxi and Didi Chuxing
- ⇒ Amazon, Alibaba, Tencent, JD.com
- ⇒ Web Blog Analysis
- ⇒ Spam Email Detection
- Online Electronic Medical Data Analysis
- Electronic Health Care and eHealth
- Social Network Analysis

### **How Amazon Makes Money**



NOW GO INNOVATE

GARYFOX.CO

# **Examples of Web Search Engines**



# **Examples of Web**



### Some Internet Related Research Projects for Students

- "Searching and Analyzing Big Data: Context-sensitive and Task-aware Approaches". This project is supported by NSERC Individual Discovery Grant (2020 - 2026)
- "Analyzing and Searching Medical Data for Cost Effective Health Care". This project is supported by Early Researcher Award/Premier's Research Excellence Award
- **"Finding Best Evidence for Evidence-based Best Practice Recommendations in Health Care". This project is supported by NSERC Collaborative R&D Grant**
- Other research projects will also be available such as IBM SUR and OCE projects etc.
- Advanced Data Mining and Machine Learning Technologies for Next Generation eHealth Decision Support Systems



 $\diamond$ 

 $\blacklozenge$ 

 $\blacklozenge$ 



# **Course Objectives**

This course will cover both programming aspects of internet applications and advanced topics of Web technology, such as information retrieval, Web search and Web mining.

The major objective of this course is for the students to learn how to analyze, design and implement internet applications and obtain a solid grasp of how techniques in Web technology can be applied to solve problems.

# **Course Content**

Introduction to client-server systems, WWW and Web technology HTML, XHTML and XML Servlets and Java Server Pages (JSP) JDBC and Java Script **E-business and E-commerce** Web mining, Web server logs and Web session Information Retrieval (indexing, search and ranking) Web search and Web crawler

# **Recommended Reading**

Internet and World Wide Web: How to Program (2nd edition)

- Deitel M., Deitel J., Nieto T. R.
- Prentice Hall, 2002

Advanced Java 2 Platform: How to Program

- Deitel M., Deitel J., Santry S. E.
- Prentice Hall, 2002

Foundations of Web Technology

Sarukkai R.

 $\bullet$ 

Kluwer Academic Publishers, 2003

Mining the World Wide Web: An Information Search Approach

- Chang G., Healey M. J. et al.
- Kluwer Academic Publishers, 2003

# **Evaluation and Grading**

3 assignments(due through the term)> Assignment 1:will be announced soon(25%)> Assignment 2:TBA(25%)> Assignment 3:TBA(20%)

**Final Exam** 

 $\diamond$ 

(**30%**) Dec

# **Group Assignments**

Each group consists of 4 or 5 members.
 There is a team leader in each group.
 The team leader should send me an email telling me how many people are in the group, their names, student Ids, email addresses and course section ID as soon as possible (before Sept 17).
 Your group assignments will be evaluated in

Your group assignments will be evaluated in terms of your programs, assignment reports and class presentations.

# **Class Time and Office Hours**

Class time: **Wednesday 19:00pm – 22:00pm Class location: Virtual Zoom Room** Office hours (in Virtual Zoom Room) By appointment **TA name: a PhD candidate** ♦ TA office hours: to be arranged

# Introduction to Client-Server Systems, WWW and Web Technology



# Outline

♦ The Internet The Web What makes the Web work? > HTTP > URL > HTML CGI Example of a Web page Summary

# **The Internet**

То

### IP Address: 123.21.12.131



# **The Internet**

### Worldwide collection of interconnected networks.

 Began in late '60s in ARPANET, a US project, investigating how to build networks that could withstand partial outages.

Starting with a few nodes, Internet estimated to have over 100 million users in 1997, and over 270 million users in over 100 countries in 1998, with one million new users joining each month.

# **Historical View: Internet**

◆ 1969 - Telnet ♦ 1970 - 4 computers > Stanford, UCLA, UC Santa Barbara, U Utah ♦ 1971 - FTP ♦ 1983 - 562 computers on the internet ♦ 1993 - 1.2 million computers on the internet ♦ 1999 - ssh, sftp, ..... ♦ 2010 - Amazon, Alibaba, .....

♦ 2020 - Smart-based devices, .....



♦ The Internet ♦ The Web What makes the Web work? > HTTP > URL > HTML > CGI Example of a Web page  $\diamond$ Summary

# The Web

### World-Wide Web (Web, WWW)

networked information system that provides a simple way of browsing different types (text, pictures, video, audio, etc.) of information on the Internet using hyperlinks.

### Web pages

electronic documents that typically contains several types of information accessible via the World Wide Web

### Web sites

a collection of related Web pages of a certain individual, group, or organization.

The Web uses a client/server model

# **Client-Server Model**

Browser - software to interact with internet data at the client

machine that services internet request



machine that initiates internet request

# **Client/Server Interaction**



# What is a Web Server?

### Web server

 $\blacklozenge$ 

- computer running application software that listens and responds to a client computer's request made through a web browser
- machine that hosts web pages and other web documents
  - provides web documents and other online services using HTTP

# What is a Web Browser?

### Web browser

- application software that is used to locate and issue a request for the page on the web server that hosts the document
- It also interpret the page sent back by the web server and display it on the monitor of the client computer computer program that lets you view and explore information on the World Wide Web

# Web Browsers







- Microsoft Internet Explorer browser integrated with the Windows operating system. Mac versions are available.
  - Netscape Navigator available for Windows, Mac, and Unix platforms.
  - **Opera** one of the alternatives to the two most popular browser mentioned above
  - Mozilla open source web browser software
  - Lynx popular Unix text-based browser
  - Google Chrome is a Google browser that combines a minimal design with sophisticated technology.

E Internet Explorer





♦ The Internet ♦ The Web What makes the Web work? ► HTTP > URL > HTML ▷ CGI … Example of a Web page Summary

# What Makes the Web Work?

### The Web relies on these mechanisms:

- <u>Protocols</u> set of standards used to access resources via the Web
- <u>Universal Resource Locator (URL)</u> uniform naming scheme for Internet resources
  - <u>HTML</u> Document formatting language used to design most Web pages
    - CGI Common Gateway Interface
  - Servlet Application run by a server connected to the WWW. It is one of the most popular avenues for Java development today.

# Protocols

- Standard set of rules that governs how computers communicate with each other, i.e. HTTP, FTP and SMTP.
- HTTP (HyperText Transfer Protocol) is the underlying protocol used to transmit information over the Web.

HTTP is based on request-response paradigm:

- Connection: Client establishes connection with Web server.
- Request: Client sends request to Web server.
- Response: Web server sends response (HTML document) to client.
- Close: Connection closed by Web server.

# HTTP Connection





> makes an HTTP request for a web page
> makes a TCP/IP connection
2. Server accepts request
> sends page as HTTP
3. Client downloads page
4. Server breaks the connection

# Uniform Resource Locators (URLs) Identifies the file to request Specifies server and file Defaults used for missing values



# protocol host computer directory path

### Note: Not all URLs will have the directory and filename



# HyperText Markup Language (HTML)

### Hypertext

 presents and relates information as hyperlinked documents that point to other documents or resources.

### HTML

- A standard markup language that defines a hypertext document.
- A simple, powerful, platform-independent document language.
- Specifies what displays should look like
- Browser interprets HTML
- Same HTML file often looks different across browsers
- > HTML files are the source files of Web pages

# **HTML File Structure**

# What About Graphics?

An HTML file can refer to an image file

# Here is a nice picture: <IMG SRC="stars.gif">

# What About Hyperlinks?

An HTML file can refer to another HTML file

<h2>Teaching</h2>

<a href="http://ai.uwaterloo.ca/<u>3421.html</u>"> COSC 3421 Fall 2002</a>

<a href="http://ai.uwaterloo.ca/3221.html"> COSC 3221 Winter 2003</a>

# **Simple Formatting**

<H1><FONT COLOR="#b80000">
Heading level 1</FONT></H1>
<H2><FONT COLOR="#ff0000">
Heading level 2 </FONT> </H2>
<P>Paragraph with <B>bold</B> and
<I>italic</I> text.</P>

<HR>



# **Creating HTML Files**

### Text editor (Notepad, Pico)

### HTML Editor (FrontPage, Netscape Gold and HoTMetaL)

# **Moving Files to Servers**

**Browser** 

Server

User views files in browser

Author sends files to server Author

creates

Author

# **Client-Server Systems Architecture**



Web page: document written in HTML, JSP and ASP.



@ 2009 Microsoft | Privacy | Legal | Help | Feedback



© 2009 Microsoft | Privacy | Legal | Help | Feedback



**Internet Banking** 

### **Uber Eats Three-Sided Marketplace Business Model**

Uber Eats is a three-sided marketplace connecting a driver, a restaurant owner and a customer with Uber Eats platform at the center. The three-sided marketplace moves around three players: Restaurants pay commission on the orders to Uber Eats; Customers pay the small delivery charges, and at times, cancellation fee; Drivers earn through making reliable deliveries on time.



Clip slide

Presentation2 - uber



### Ali ECOSYSTEM of E/M/S COMMERCE





China's E-commerce Empire: Alibaba Group



# **Wechat Functions**

**Functions within WeChat** 



# Wechat Business Model



Instant messaging



### Social networking



Gaming



Payment



Smart home services



Government services



Third-party services





Shopping

### THE AMAZON BUSINESS MODEL

(simple version)



NOW GO INNOVATE

GARYFOX.CO

# **Static and Dynamic Web Pages**

A static Web page is ready before it is accessed.

The content of a dynamic Web page is generated each time it is accessed.

# **Common Gateway Interface** (CGI)

 CGI programming techniques were introduced to provide dynamic Web pages via server-side interaction.

 A standard method to extend the functionality of the web server.

**♦** 

Any programming language can be used. Common ones include: Perl, C++, Visual Basic.

# **CGI-based Web Application**



# **How Web Page Works**

html> head>

### Sample web page and its source.

- The source contains the instructions that define the contents, layout, and structure of a web page.
  - The instructions are written in HTML or another web authoring tool used in creating the page. The browser uses these instructions to interpret and

display the web page on the screen.



# **How Web Page Works**

- 🗆 × ICT for Teacher Training Portal - Microsoft Internet Explorer - [Working Offline] File Edit View Favorites Tools Help 0 3 4 J 3 \* W URL Back Stop Refresh Home Search Favorites Media History Mail Print Edit Discuss Forward -2 GO Address 🍘 http://www.unescobkk.org/ips/ict/ict.htm LINKS @ UP WebMail @ Welcome to Eudora Web-Mail @ Welcome to AmExMail 2 Ramon's Excite Start Page 2 Sign in - Yahoo! Groups **Navigational tools** Knowledge Resources Virtual Library Webcast Portals Electronic Publications Search of the Month **Electronic Articles** Database sections ICT Main Page Photo Library Ordering Publications ▶ ICT in Education search Teachers' Roles in the I Environment The Information and Communication Technology ICT Strategies and Onlin Navigational Portal (ICT) for Teacher Training is the gateway to Courses Graphics / Internet resources and websites dedicated to Integrating ICT into training the teachers in utilizing information and Teaching for teacher training communication technologies to enhance their **Hyperlinks** Teaching Ideas, Lessons teaching skills. and Curriculum Materials Educational Related Topics: To bridge the digital divide in-Software/Courseware region, the Japanese Government through its Japanese Funds-in-Trust mechanism is providing substantial Using Internet Resource to UNESCO, Bangkok to develop, oversee, and implement a project that will bring ICT to education, operifically, the project is aimed at contributing to bridging the digital divide by exploring and demonstrating bern new ICT can Electronic Collaboration be used to reduce disparities in educational access and quality and bu providing corresponding access to knowledge, new educational methods, learning materials and enturally relevant contact in schools and lear Bringing Your Classroom Online places in marginalized communities in selected countries in the region Evaluation Tools and Indicators **Hyperlinks 3**<sub>101</sub> IC in Education 🖌 Тр ners' Roles in the ICT Training Strategie Links to articles, researd Enviror O hae ling Cources studies and papers which The first part describes Websites in this section point explain how education out the new knowledge, skills experiences and has charged to adapt to and competencies required o strategies on the new knowledge society and how professional teachers in an ICT nformation technology has environment as well as explain the development on ICT for teachers ontributed to the educational changes that have been brought about and the second part provides links reform and improvements in in the teaching/learning environment to numerous online courses for teaching and learning. due to the introduction of ICT and how upgrading teachers' skills in the www.unescobkk.or use of ICT in teaching. they impact on the roles of teachers. 🙁 Internet

# Cookies

♦ A piece of information generated by the web-server and stored in the client side ready for future access. Cookies can make CGI scripts more interactive. Cookies are text files stored on Web client. CGI script creates cookie and has a Web server sent it to client's browser to store on hard disk. Later, when client revisits Web site and uses a CGI script that requests this cookie, client's browser sends information stored in the cookie.

# Cookies

### How do cookies work?



## Where are cookies used?

- Shopping applications
- Storing login information
- Tracking pages visited by a user



The Web is a networked information system that contains a huge collection of files

- The Web relies on clients and servers
  - HTML and other files are sent from servers to clients
  - Files are identified by URLs
    - Servers send files to browsers
    - Browsers interpret HTML
  - Cookies is a piece of information generated by the web-server and stored in the client side.