

Lecture 17

World Wide Web



You might like...

- FRS 116 The Evolution of Human Language, Christiane D. Fellbaum
- FRS 122 The Digital World: Connection and Communication, Swati Bhatt
- FRS 134 Scientists Against Time, Harold Feiveson
- FRS 159 Science, Technology, and Public Policy, Harold Shapiro
- COS 126 Computer Science: An Interdisciplinary Approach
- SOC 409/COS 409 Critical Approaches to Human Computer Interaction, Janet Vertesi
- STC 109 Transformations in Engineering and the Arts, Adam Finkelstein, Jeff Snyder, ...
- HUM 346/ENG 256 Introduction to Digital Humanities, Emily McGinn
- JRN 260 The Media in America: What to Read and Believe in the Digital Age ,
Joe Stephens
- SOC 277 Technology and Society, David Reinecke

The World Wide Web

- a brief history
- how it works
- cookies, Javascript, other mechanisms
- advertising, tracking, surveillance
- how to defend yourself: privacy and security
- technical issues
- political / legal / social / economic / jurisdictional issues

(World Wide) Web

- a way to connect computers that provide information (servers) with computers that ask for it (clients like you and me)
 - uses the Internet, but it's not the same as the Internet
- **URL** (uniform resource locator, e.g., <http://www.amazon.com>)
 - a way to specify what information to find, and where
- **HTTP** (hypertext transfer protocol)
 - a way to request specific information from a server and get it back
- **HTML** (hypertext markup language)
 - a language for describing information for display
- **browser** (Firefox, Safari, Chrome, Edge, ...)
 - a program for making requests, and displaying results
- **embellishments**
 - pictures, sounds, movies, ...
 - loadable software
- **the set of everything this provides**

Web history

- **1989: Tim Berners-Lee at CERN**
 - a way to make physics literature and research results accessible on the Internet
- **1991: first software distributions**
- **Feb 1993: Mosaic browser**
 - Marc Andreessen at NCSA (Univ of Illinois)
- **Mar 1994: Netscape**
 - first commercial browser
- **technical evolution managed by World Wide Web Consortium**
 - non-profit organization at MIT, Berners-Lee is director
 - official definition of HTML and other web specifications
 - see www.w3.org



HTTP: Hypertext transfer protocol

- What happens when you click on a URL?
- client opens TCP/IP connection to host, sends request

```
GET /filename HTTP/1.0
```

- server returns
 - header info
 - HTML



- server returns text, which can be dynamically created as needed
 - can contain encoded material for images, music, video (MIME format)

- URL format

```
service://hostname/filename?other_stuff
```

- *filename?other_stuff* part can encode
 - data values from client (forms)
 - request to run a program on server (cgi-bin)
 - anything else

Embellishments

- original design of HTTP just returns text to be displayed
- **MIME** format for pictures, sound, video, ...
 - helpers or plug-ins display non-text content
- **forms** filled in by user
 - needs a program on the server to interpret the information (cgi-bin)
- **cookies** to remember information on client
 - HTTP is stateless: server doesn't save anything from one request to next
 - cookies are a way to remember information at the client
- **Javascript**: download code to run on the client

Forms and CGI programs

- **"common gateway interface"**
 - standard way to request the server to run a program
 - using information provided by the client via a form
- **if the target file on the server is an executable program**
- **and if it has the right properties and permissions**
 - e.g., in /cgi-bin directory and executable
- **then run it on the server to produce HTML to send back to the client**
 - using the contents of the form as input
 - output depends on client request: created on the fly, not just a file
- **CGI programs can be written in any programming language**
 - e.g., Python, Java, ...

Cookies

- HTTP is stateless: it doesn't remember from one request to the next
- cookies are intended to deal with stateless nature of HTTP
 - remember preferences, manage "shopping cart", etc.
- **cookie: one chunk of text sent by server to be stored on the client**
 - stored in browser while it is running (transient)
 - stored in client file system when browser terminates (persistent)
- **when the client reconnects to same domain,**
 - browser sends the cookie back to the server**
 - sent back verbatim; nothing added
 - sent back only to the same domain that sent it originally
 - contains no information that didn't originate with the server
- in principle, pretty benign
- but pervasively used to monitor browsing, for commercial purposes

Cookies are not the only tracking mechanism

- 3th party cookies are decreasing in value
as more browsers block them by default
- Alternatives:
- JavaScript
 - potentially continuous monitoring and reporting of activity on a page
- web bugs, web beacons, single-pixel gifs
 - tiny images that report the use of a particular page
 - these can be used in mail messages, not just browsers
- HTML canvas fingerprinting
 - uses subtle differences in browser behavior to distinguish users

Javascript

- programming language loosely in the C family (surface syntax similar)
 - (no relationship to Java)
 - compiled into instructions for a virtual machine
 - like the Toy machine on steroids
 - instructions are interpreted by a virtual machine in browser
- most common use is embedded in web pages, running in browser
 - can also run standalone

```
<script> ... </script>
```
- can interact with browser to see what is displayed, change what is displayed
 - can watch events like clicks, mouse motion, ...
 - can send and receive data from network (with restrictions)
 - can load more Javascript from network (with restrictions)

What does Facebook know about you?

- <https://www.cnn.com/2018/03/27/facebook-knows-a-lot-about-me.html>
- It can recognize my face
- It knows every ad topic I've ever clicked
- It has a list of every company that has my contact information from the ads I've clicked
- It has a list of every contact in my phone book
- It knows every social event I was invited to and/or attended through Facebook
- It has a log of every friend I have on Facebook and when we became friends
- It knows every time I logged in
- It has a copy of my timeline going back to the time I joined
- It knows my major life events
- It knows every video I've watched on Facebook
- It knows exactly where I was
- It has old messages
- It has a copy of every photo I've ever uploaded