Web Technologies

Topics

- How does the Web Work
- Evolution of the Web
  ◦ Enable transactions
  ◦ Allow interactivity between browser and server
  ◦ Facilitate personalization
  ◦ Support multiple browsing devices
  ◦ Better organize and retrieve Web content
  ◦ Support Business-to-Business applications
  ◦ IoT (Internet of Things)

World Wide Web
The Triumph of Anarchy

- Perhaps the most important human technological artifact that evolved more or less ad-hoc
- Limited original vision of the WWW has very little to do with today's impressive reality
- Web Users have consistently innovated in figuring out new ways of leveraging this powerful medium
- Web architects then try to catch up by extending (read "patching") the Web infrastructure to support these new uses

---

How it all started…
The Web as a Static Document Repository

- Tim Berners-Lee's original vision for the WWW (1989)
- An easy way to access cross-linked static documents stored in a variety of servers around the world
- Initial specification defined:
  ◦ A language for formatting such documents (HTML)
  ◦ A simple protocol for communicating between browsers and servers (HTTP)

The turning point

- Use before MOSAIC: a text based email, bulletin boards (gopher), ftp, telnet
- 1993 -Marc Andreesen (student at UIUC) writes Mosaic first graphical WWW browser
- precursor of Netscape

How the (original) Web works

Anatomy of a URL

URL = Uniform Resource Locator

http://web.mit.edu/sloan/www/index.html

Name of protocol for communication with server (http is standard web protocol)
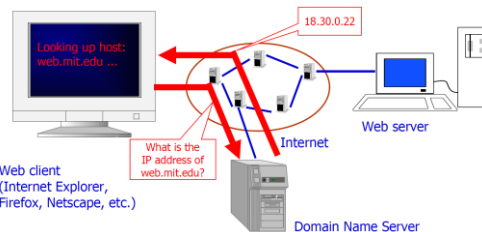
Domain name of web server where page is stored

Pathname of file within web server's local file system

- index.html : default page to display
- On the Server: There is a default root folder for the website
  ◦ /var/www/vhosts/pinformatics.tamhsc.edu/httpdocs
    ▪ Location of : pinformatics.tamhsc.edu/ or pinformatics.org
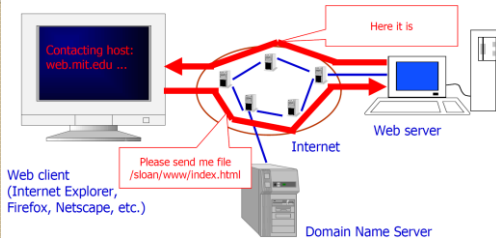 ◦ Specified in the webserver software (e.g. Apache)

How the Web Works (1)

Open Location: http://web.mit.edu/sloan/www/index.html

Looking up host: web.mit.edu ...

18.30.0.22

Web client (Internet Explorer, Firefox, Netscape, etc.)

What is the IP address of web.mit.edu?

Internet

Web server

Domain Name Server

How the Web Works (2)

Open Location: http://web.mit.edu/sloan/www/index.html

Contacting host: web.mit.edu ...

Here it is

Please send me file /sloan/www/index.html

Web client (Internet Explorer, Firefox, Netscape, etc.)

Internet

Web server

Domain Name Server

How the Web Works (3)

Open Location: http://web.mit.edu/sloan/www/index.html

Web client (Internet Explorer, Firefox, Netscape, etc.)

Internet

Web server

Domain Name Server
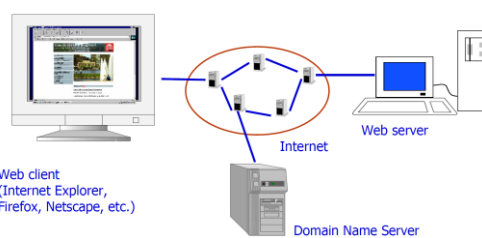
Web is a Client/Server System

- Web Clients
  ◦ Use HTTP protocol to connect to servers
  ◦ Request and display Web pages stored in servers
  ◦ Typical clients: Web browsers
- Web Servers
  ◦ Listen for incoming connections from clients
  ◦ Use HTTP protocol to converse with clients
  ◦ Store and transmit Web pages to clients
  ◦ Example: Apache Web server (38%), MS (33%)

Looking at a Web Page

- User request from a browser a page via URL
- Brower asks DNS for the IP address of the MIT Server using the domain name of URL
- DNS replies with 18.170.0.167
- Browser opens TCP connection to 18.170.0.167
- Browser sends the command
  ◦ GET/class/syllabus.htm
  ◦ file path and HTTP method used
- The MIT Server sends file syllabus.htm
- TCP connection is released
- Browser displays the contents of syllabus.htm

Topics

- How does the Web Work
- Evolution of the Web
  ◦ Enable transactions
  ◦ Allow interactivity between browser and server
  ◦ Facilitate personalization
  ◦ Support multiple browsing devices
  ◦ Better organize and retrieve Web content
  ◦ Support Business-to-Business applications
  ◦ IoT (Internet of Things)

Summary:
WWW The Original Concept (1989-1995)

- Connect Human Readers to Interconnected Static Documents
- Main advantage
  ◦ Universality
- Main disadvantage
  ◦ Lack of interactivity
  ◦ Yet another mass broadcast medium
- Main business use
  ◦ Awareness building

Summary: Web
The Current Concept (1995 – today)
2005? 2015?

- Connect Human Users to Documents and Applications
- Main business use
  ◦ B2C (business 2 customer) Transactions
  ◦ Customer Support

Summary: Semantic Web
Supporting Technology (today – future)
2005? 2015?

- Semantic interoperability among independently-developed Web applications
- Main business use
  ◦ B2B (business 2 business) Transactions

Summary: Evolution of The Web
Acknowledgement: Thomas Malone, Chris Dellarocas

Now?

Presence	Transactions	Business
Publish Information	**Process Transactions**	**Digital Economy**
Collection of Static Documents	Collection of Dynamic Pages (DP)	Social Media, IoT Semantic Web
Websites	Wed-enabled Existing Systems	Business Transformation
Pages	Transactions	Business Processes
Islands	Islands	Constellations
Eyeballs	Revenue	Profits

Evolution of the Web

- Business Drivers
  - Enable transactions
  - Allow interactivity between browser and server
  - Facilitate personalization
  - Support multiple browsing devices
  - Better organize and retrieve Web content
  - Support Business-to-Business applications
  - IoT (Internet of Things)

The Web as a transaction facilitator

- Business Motivation: Low-cost front-end for allowing customers to connect to corporate computers
  - Customer registration/Address changes
  - Order tracking/Customer support
  - Online Transactions: eCommerce
- Problems of original Web concept
  - Static web pages
  - No interactivity
  - Stateless protocol: no support for multi-step transactions
 - Insecure communications

Web Forms

- Pages which contain fields to be filled by user
- Usually contain a "Submit" button
- When user presses "Submit", server responds by sending a page containing information specific to the user-supplied parameters
- Examples:
  - Web search tools
  - Order forms in commercial web sites

Web Forms Under the Hood

- Server sends original html page containing input fields
- User types info into fields and presses submit button
- Client establishes connection with handler script at server side (script filename contained in web page)
- Client collects user input into a long string and sends it along with an HTTP command back to server
  - POST customer=John+Doe&cardno=1234567890 &expires=6/98&product=123&quantity=5
- Handler script at server reads parameter string and processes it, usually producing a new page as a result
  - Scripts typically translate parameters into SQL statements for a database and translate the query results into an HTML page
  - Example: Common Gateway Interface (CGI), Microsoft Active Server Pages (ASP), Coldfusion, PHP, Python, Rudy, etc

Example: FedEx
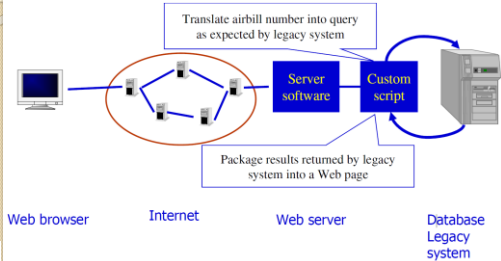
Evolution of the Web

- Business Drivers
  - Enable transactions
  - Allow interactivity between browser and server
  - Facilitate personalization
  - Support multiple browsing devices
  - Better organize and retrieve Web content
  - Support Business-to-Business applications
 - IoT (Internet of Things)

The Interactive Web

- Business Motivation:
  - Allow complex interaction between user browser and corporate server
  - Web becomes an extension of the user's PC
  - Browser becomes a window to a variety of corporate applications
- Problems with Web Forms/CGI/ASP
  - All processing done at server side
  - Rapid user interaction with Web page not possible
  - Need local processing to create highly interactive Web pages

Enter Java

- Client side software: Programming language to enable interactive Web pages
- Developed by Sun Microsystems
  - originally for programming intelligent microwave ovens!!!
- Initially, Java applets
  - Java executable software
  - Applets are platform-independent (run equally well on Windows, Macs, Unix, etc.)
  - Require special browsers that can support Java
- Now Javascript
  - Human readable programs, interpreted on client

Java Applets: Not used much now
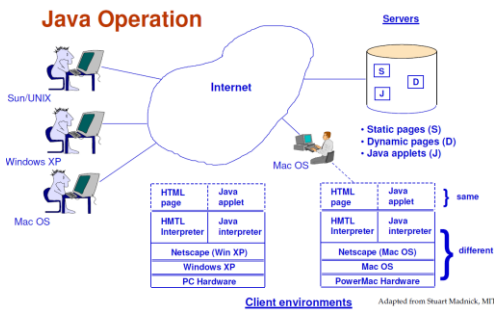
Java Operation

Evolution of the Web

- Business Drivers
  - Enable transactions
  - Allow interactivity between browser and server
  - Facilitate personalization
  - Support multiple browsing devices
  - Better organize and retrieve Web content
  - Support Business-to-Business applications
  - IoT (Internet of Things)

Personalized Interaction

- Business motivation:
  - Low cost medium for gathering information from customers to allow
    - Personalized service
    - Targeted advertising
- Problems with current model
  - Does not allow easy identification of distinct customers

Cookies

- A method for identifying web users and delivering customized web sites
  - First time user connects to a web site, s/he is asked to fill in personal information form
  - Server packages information into a "cookie" file and sends cookie to browser
  - Browser stores cookie in local file system
  - Each subsequent time browser visits site, it sends cookie back to server
  - Server uses information stored in cookie to identify user and possibly customize the supplied web pages
- Privacy & Security implications?
  - Monitor without permission: fine grain cookie permission control by site

The user's browser passes the contents of the cookie to the Web server. The server uses this cookie to look up information about the user (this information may have been gained from a previous registration process). The information (specifically, the user's age) is used to query a database via ODBC to determine which products should be recommended to this user. The resulting list of products is inserted into a dynamically created HTML page and sent to the user's browser for display.

Figure by MIT OCW

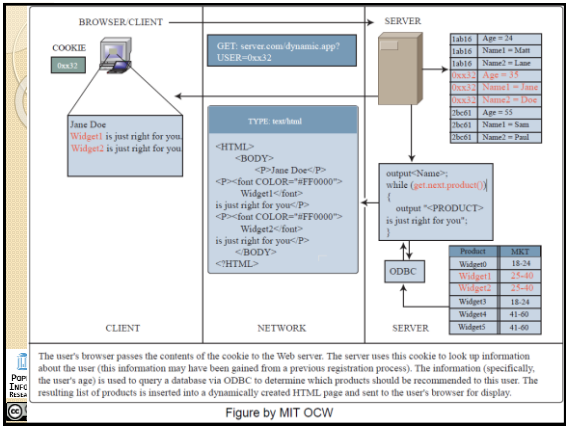## Evolution of the Web (Presented out of order)

- Business Drivers
  - Enable transactions
  - Allow interactivity between browser and server
  - Facilitate personalization
  - Support multiple browsing devices
  - Better organize and retrieve Web content
  - Support Business-to-Business applications
  - IoT (Internet of Things)

Support of B2B Applications

- Original Web was conceived as a communication medium between computers and humans
- Amazing new applications will become possible if computers can automatically read and understand Web pages
  - Electronic purchasing
  - Intelligence gathering
  - ....
- Problem:
  - HTML pages are unstructured
  - HTML only provides information about presentation, not meaning (semantics for computers)

What is the underlying issue?

- When storing documents on the web, specify not only their appearance, but also their semantics (i.e. their meaning for computers to understand!)

The Semantic Web I

- The "Next Generation Web" with well-established infrastructure for expressing information in a
  - Precise, Human-readable, and Machine-interpretable form.
  - The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network.
- Enable syntactic and semantic interoperability among independently-developed Web applications, allowing them to efficiently perform sophisticated tasks for humans.
- Enable Web resources to be accessible by their semantics rather than by keywords and syntactic forms.
- Enable inferences:
  - Chris is an associate professor
  - Associate professors are permanent employees
  - Chris is a permanent employee (inference)

The Semantic Web II
http://www.w3.org/standards/semanticweb/

- Vocabulary management: Ontology
  ◦ OWL: to build ontology
  ◦ SKOS (Simple KOS): for designing Knowledge Organization Systems (KOS)
- Linked Data
  ◦ Entity & Relationships: "subject-predicate-object" triples
    · Tamoxifen treats malignant neoplasm of breast
- Query language for Semantic Web: SPARQL
  ◦ Semantic web ~ global database
- Rule based Inference: reasoning over data
  ◦ Tamoxifen treats malignant neoplasm of breast
  ◦ Susan has malignant neoplasm of breast
  ◦ INFERENCE: Tamoxifen can treat Susan
- Vertical Application
  ◦ By industry (e.g. health care)

Example: Semantic Web

- Comparison of prices on different sites
  ◦ May not be implemented with formal semantic web technology
  ◦ Aggregating website: directly communicating with individual websites

$490 round trip	2:16 PM – 10:58 PM	10h 42m	2 stops
	Delta · Skywest DBA Delta Conn		MSP, LAX
$538 round trip	11:35 AM – 10:25 PM	12h 50m	2 stops
	American · Skywest Airlines as A · IAH-OAK		ORD, PHX
$555 round trip	1:29 PM – 10:25 PM	10h 56m	2 stops
	American · Skywest Airlines as A · IAH-OAK		ORD, PHX
$640 round trip	3:35 PM – 10:58 PM	9h 23m	1 stop
	United, Delta · Compass DBA De · IAH-OAK		4h 08m in LAX
$839 round trip	5:45 PM – 11:29 PM	7h 44m	1 stop
	Alaska · Skywest Airlines as Alas · IAH-OAK		41m in SEA

Ontology: common vocabulary

- Manage vocabularies and relationships to represent knowledge
- Taxonomy: Hierarchical relationships only
- Examples
  ◦ Health Care Provider Taxonomy Codes
    · a hierarchical code set that consists of codes, descriptions, and definitions
    · designed to categorize the type, classification, and/or specialization of health care providers
    · National Uniform Claim Committee (NUCC), AMA
    · published (released) twice a year on July 1st and January 1st.
    · The July publication is effective for use on October 1st and the January publication is effective for use on April 1st.
    · http://www.nucc.org/index.php?option=com_content&view=article&id=14&Itemid=125
  ◦ ICD 9, ICD 10
  ◦ CPT code, HPCPS (Healthcare Common Procedure Coding System) code
  ◦ SNOMED clinical terms: Unified Medical Language System (UMLS)

Linked Data
http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/

- Entity & Relationships
  ◦ Entities
    · Max
    · Kiel
    · Germany
  ◦ Relationships
    · As Triples
    · M born in K
    · K located in G
- Datasets
  ◦ Dbpedia
  ◦ FOAF
    · Friend of a Friend

Triples: subject-predicate-object
http://skr3.nlm.nih.gov/SemMed/SemMedQuickTour/SemMedQuickTour_player.html

- The entire database is a set of "subject-predicate-object" triples

Query language for Semantic Web SPARQL

- SPARQL query example
- "What are all the country capitals in Africa?":
- In this query, the prefix "abc" stands for "http://example.com/exampleOntology#".
- Variables are indicated by a "?" or "$" prefix. Bindings for ?capital and the ?country will be returned.
- The SPARQL query processor will search for sets of triples that match these four triple patterns, binding the variables in the query to the corresponding parts of each triple.

```
PREFIX abc: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
  ?x abc:cityname ?capital ;
     abc:isCapitalOf ?y .
  ?y abc:countryname ?country ;
     abc:isInContinent abc:Africa .
}
```

Rule Based Inference:
Reasoning over data through rules

- Example 1
  - Tamoxifen treats malignant neoplasm of breast
  - Susan has malignant neoplasm of breast
  - INFERENCE
    - Tamoxifen can treat Susan
- Example 2
  - Chris is an associate professor
  - Associate professors are permanent employees
  - INFERENCE
    - Chris is a permanent employee

Vertical Application
Semantic Web Health Care and Life Sciences (HCLS) Interest Group

- Develop Applications by industry
- Example: NPI (National Provider ID)
  - the National Plan & Provider Enumeration System (NPPES) at CMS
  - The NPI (National Provider Identifier) number is a unique 10-digit number issued by the CMS to healthcare providers in the United States.
  - The NPI is the only healthcare provider identifier that can be used for identification purposes in standard transactions by covered entities
  - Publish Open NPI DB (not triples), weekly updates
  - http://www.bloomapi.com/
    - Makes it available as API (for other computers to access)
  - https://npidb.org/
  - http://docnpi.com/

Break
Read Over and do Lab 4 Discussion

Remainder of Class

- 10 min break
- Lab 4
  - Read (5 min)
  - Discussion (15 min)
  - Write (10min)
    - You MUST list the discussion group members in your essay
- Lec 2

Evolution of the Web

- Business Drivers
  - Enable transactions
  - Allow interactivity between browser and server
  - Facilitate personalization
  - Support multiple browsing devices
  - Better organize and retrieve Web content
  - Support Business-to-Business applications
  - IoT (Internet of Things)

Multiple Delivery Devices

- Business motivation:
  - Allow users to access web content from a variety of devices
    - PC Browsers
    - Smart Phones
    - Telephones (via voice interface)
    - Tablets
    - PDAs (e.g. Palm Pilots)
    - …
- Problems of current Web model
  - Each access device has different look-and-feel requirements
  - HTML specifies formatting for PC browsers only

Responsive Web Design

- Adapts the layout to the viewing environment
  - The fluid grid concept calls for page element sizing to be in relative units like percentages, rather than absolute units like pixels or points.
  - Flexible images are also sized in relative units, so as to prevent them from displaying outside their containing element.
  - Media queries allow the page to use different CSS style rules based on characteristics of the device site is being displayed on, most commonly the width of the browser.
- Bootstrap.css: open source

Cascading Style Sheets (CSS)

- Used to describe the look and formatting of the document written in a markup language
- Most often used with HTML
- Designed to separate document content from document presentation (e.g., layout, colors, and fonts)
  - Power of Computer: separation
    - Modular programming
- http://www.csszengarden.com/

```
h1 { color: white;
     background: orange;
     border: 1px solid bl
     padding: 0 0 0 0;
     font-weight: bold;
}
/* begin: seaside-theme */
body {
     background-color:white;
     color:black;
     font-family:Arial,sans-serif;
     margin: 0 4px 0 0;
     border: 12px solid;
}
```

Using Media in CSS

```
<link href="css/print.css" rel="stylesheet" type="text/css"
     media="print">
<link href="css/large.css" rel="stylesheet" type="text/css"
     media="only screen and (min-width:1500px)">
<link href="css/medium.css" rel="stylesheet" type="text/css"
     media="only screen and (max-width:1499px)">
<link href="css/tablet.css" rel="stylesheet" type="text/css"
     media="only screen and (max-width:1167px)">
<link href="css/mobile.css" rel="stylesheet" type="text/css"
     media="only screen and (max-width:650px)">
```

Using Media in CSS
What does it mean?

- If printing on paper
  - use print.css
- If "on screen" and width<=650
  - Use mobile.css
- If "on screen" and 650<width<=1167
  - Use tablet.css
- If "on screen" and 1167<width<=1499
  - Use medium.css
- If "on screen" and 1499<width
  - Use large.css

HTML5

- Hyper Text Markup Language
- Commands = Tags: <TAG></TAG>
  - <html></html>
  - XML: Extensible Markup Language
    - Main issue: filesize too big
- DOM: Document Object Model
  - Object Oriented
  - Page is composed of objects
    - Title, header, paragraph, table (row <tr>, cells <td>)
  - Objects have attributes
    - Color, font, size
  - Set attributes
    - Initially: css (layout) & html
    - Dynamically change: javascript

HTML5 components

- HTML: content
  - Objects: <head><body><footer>
  - Viewable (rendered) and nonviewable objects
- CSS: layout (color, format, font etc)
  - Define *attributes* of an object
- Javascript: action = dynamically change
  - Click a button, update data etc
  - Create, delete objects
  - Dynamically change *attributes* of an object
  - AJAX: asynchronous Javascript and XML
    - Group of interrelated web development techniques used on client-side to create asynchronous web applications
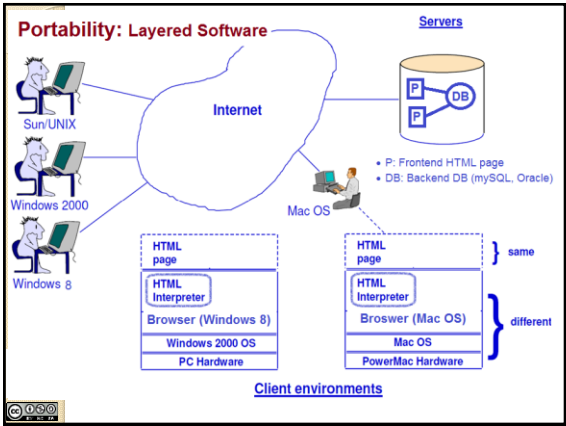
Evolution of the Web

- Business Drivers
  - Enable transactions
  - Allow interactivity between browser and server
  - Facilitate personalization
  - Support multiple browsing devices
  - Better organize and retrieve Web content (optional)
  - Support Business-to-Business applications
  - IoT (Internet of Things)

Organize and Index Web Content

- Web is useless unless we can easily locate relevant resources
- Current solution: Search Engines
  - Index the Web by automatically "discovering" web pages and organizing them around keywords found in their text

How does Google Work?

- Before you ever enter a query:
  - Programs (called "web crawlers" or "spiders") follow links from one page to another all over the web.
  - The programs construct indexes of which words appear on which pages and save the indexes (and often copies of the pages) on massive "server farms" maintained by Google.
  - Each page is also assigned a "page rank" based on the number of other pages that link to it. Links from pages that, in turn, have lots of other pages linking to them are weighted more heavily.

Problems with today's searches

- Text keywords are misleading…
- HTML does not give any clues as to the true meaning of the data
- Medium of documents for people
  - Rather than data and information that can be processed automatically by computers
- Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." Scientific American, May 2001, pp. 35-43.

Evolution of the Web

- Business Drivers
  - Enable transactions
  - Allow interactivity between browser and server
  - Facilitate personalization
  - Support multiple browsing devices
  - Better organize and retrieve Web content (optional)
  - Support Business-to-Business applications
  - IoT (Internet of Things)

IoT (Internet of Things)

- The inclusion of electronics and software in any device not usually considered computerized in nature
- To enable it to achieve greater value and service
  - by giving it an ability to network and communicate with other devices.
  - each item is uniquely identifiable through its embedded computing device
  - but is able to interoperate within the existing Internet infrastructure
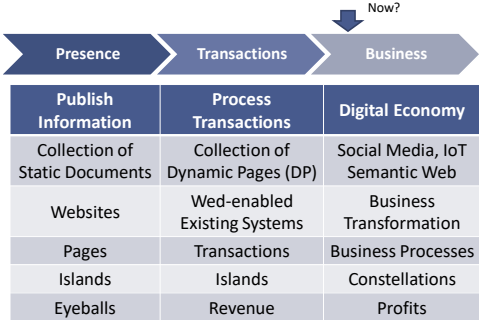- Example: home health

Take Away I: The Web

- Anatomy of a URL (Uniform Resource Locator)
  - http://pinformatics.tamhsc.edu/phpm631
- What is DNS (Domain Name Server)?
- How does the Web Work
  - User request a page via URL
  - Client uses IP address to locate the Web server
    - looks up IP from DNS if name is used
  - Client sends a request for page to the Server
    - Using the protocol specified in the URL
    - File path is specified in the URL
  - The server sends the requested page using specified protocol (http or https)
  - Client displays (renders) the page on the screen using a browser

Take Away II: Web technologies

- Static Page
- Dynamic Page (Web Forms)
- HTML5 (html, CSS, Javascript, DOM: Document Object Model)
- Cookies
- Semantic Web
  - Why: semantic interoperability among independently-developed Web applications
  - Ontology: manage vocabularies
    - knowledge representation using vocabularies and relationships
  - Linked Data
    - Entity & Relationships: "subject-predicate-object" triples
      - Tamoxifen treats malignant neoplasm of breast
  - Query language for Semantic Web: SPARQL
    - Semantic web ~ global database can be queried
  - Rule based Inference: reasoning over data
    - Tamoxifen treats malignant neoplasm of breast
    - Susan has malignant neoplasm of breast
    - INFERENCE: Tamoxifen can treat Susan
  - Vertical Application
    - By industry (e.g. health care)
- IoT

Take Away III: Evolution of The Web
Acknowledgement: Thomas Malone, Chris Dellarocas

Now?

Presence	Transactions	Business
Publish Information	**Process Transactions**	**Digital Economy**
Collection of Static Documents	Collection of Dynamic Pages (DP)	Social Media, IoT Semantic Web
Websites	Wed-enabled Existing Systems	Business Transformation
Pages	Transactions	Business Processes
Islands	Islands	Constellations
Eyeballs	Revenue	Profits

Review: Key Layers of the Internet
http://en.wikipedia.org/wiki/File:Internet_Key_Layers.png