The World Wide Web

- Relies on the Internet:
  - LAN (Local Area Network) connected via e.g., Ethernet (physical address: 00-B0-D0-3E-51-BC)
  - IP (Internet Protocol) for bridging separate physical networks (IP address: 128.100.27.199)
  - TCP (Transmission Control Protocol) for (fairly) reliably sending streams of bytes to various TCP ports on an computer on an IP network.
  - DNS (Domain Name Service) for translating names and domains into IP addresses
    (penny.cs.toronto.edu = 128.100.27.199)
  - Much more later in the course

Internet

- ARPANet online in '69
- TCP/IP selected as standard protocol in '82
- NSFNet widely deployed in '86
  - uucp
  - telnet
  - usenet
  - ftp
  - mailto

WWW

- Tim Berners-Lee at CERN proposal in '89 to have a CERN Intranet so researchers could find results of others
- Based on earlier Hypertext work dating back to '69, but especially work in '85 at Xerox (*NoteCards*) and '87 at Apple (*HyperCard*)
- A collection of protocols for moving content around
  - especially http (hypertext transport protocol)
  - maintained by the W3C (world-wide-web consortium) (http://www.w3.org)
- '93 Mosaic from NCSA: X11 on UNIX

Browsers and Servers

- Web Servers
  - Microsoft IIS, Apache, …
- Web Browsers (clients)
  - Netscape, Microsoft IE, …
- URLs (Uniform Resource Locators)
  - <scheme>:<scheme-specific-part>
  - http:<http-specific-part>
  - http://<fully-qualified-domain-name>/<path>
  - http://www.cs.toronto.edu/~penny/teaching/csc309
- MIME (Multipurpose Internet Mail Extensions)
  - <type>/<subtype>
  - text/html

Static Client Case

- A URL is typed into a client browser.
  - http://www.cs.toronto.edu
- DNS is used to translate the name into an IP address
  - 128.100.1.32
- The client browser makes a connection to TCP port 80 on the server machine and sends an HTTP GET request
- The Web server running on 'christie.cs' fetches a default page and returns it to the browser
  - The response header indicates the MIME type of the data (often 'text/html') and the character encoding (often 'iso-8859-1') and the browser sets to work displaying it
- Server may have it's way with your Web page before sending it on
  - e.g., forwarding

Try it!

```
telnet www.cs.toronto.edu 80
    Trying 128.100.1.32…
    Connected to christie.cs.toronto.edu.
    Escape character is '^]'.
GET /. HTTP/1.0 \n\n
    HTTP/1.1 200 OK
    Date: Mon, 10 Sep 2001 16:08:47 GMT
    …
    Content-Length: 3296
    Content-Type: test/html

    <html>
    …
    </html>
    Connection closed by foreign host.
```

Dynamic Client Case

- Same as the above, but the page contains references to code that executes within the browser
  - JavaScript
  - Java Applets
  - Flash plugin
- Security is always an issue with dynamic client content.
  - keeping the client machine safe from intruders

Dynamically Served Content

- Requests for certain pages result in programs being run on the server side.
  - These programs generate information that then gets returned to the browser (usually text/html).
- Common technologies
  - CGI with PERL
    - "common gateway interface"
    - HTTP GET/POST methods used to send client-entered data to the server program
  - Server-side includes
  - ASP, JSP
  - Java Servlets
- Security also an issue, but the other way:
  - keeping the server machine safe from intruders

N-Tiered Web Architectures

- The main purpose of dynamically served content are the side-effects
  - access to e-mail
  - access to a local file system
  - access to a database
- Can also have access to a custom-written server program.
  - "3-tiered" or "n-tiered" architecture

Compatibility

- Backwards Compatible
  - Old documents work on new browsers
    - must continue to allow deprecated features
    - must continue to support older syntax
- Forwards Compatible
  - Old browsers can render newer documents
    - requires hacks in the document
- Sideways Compatible
  - Different browsers can render the same document
    - need for standards

Coping

- Standards are evolving rapidly
- When designing Web pages, one must always be aware of the lowest common denominator and make the pages act reasonably for them.
- One should not design for the latest and greatest
  - i..e, IE5.5 1024x768 32-bit colour
  - unless for use in a controlled Intranet environment
- See http://www.webreview.com/browsers
- In this course, for simplicity, you may assume a fixed environment (that on CDF) as for a corporate intranet.