COMPSCI 111 / 111G

Mastering Cyberspace:
An introduction to practical computing

World Wide Web

Hypertext

Hypertext
- Text with links
- Allows free association
- More creative learning

Vannevar Bush
- 1945 published idea of hypertext using MEMEX system

Ted Nelson
- 1960 Xanadu project - first computer hypertext system
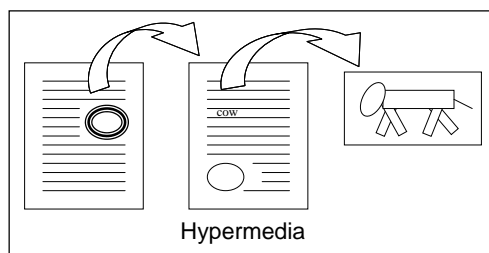
Tim Berners-Lee
- 1989 starts the WWW project at CERN

http://en.wikipedia.org/wiki/Www

Multimedia and Hypermedia

Multimedia
- The integration of many forms of media
- Text
- Images
- Sound
- Animation

cow

Hypermedia

Hypermedia
- The combination of Hypertext and Multimedia
- Links are made between any media (Hyperlinks)

The WWW project

Background: CERN
- Many networks existed
- Each network had many documents

Aims
- Access documents from any network in seamless manner
- World-Wide (distributed)
- Easy to add documents (dynamic)

Proposal
- Use Hypertext
- No intention to support hypermedia
- Research only

http://en.wikipedia.org/wiki/WorldWideWeb

Rapid Evolution

1989	Tim Berners-Lee begins work on the WWW project
1991	WWW operational at CERN
1992	WWW goes public
1993	Mosaic created by Marc Andreessen (First GUI browser)
1994	US Senate allow commerce on Internet Netscape Communications formed, Yahoo! formed
1995	Microsoft Internet Explorer
1998	Netscape became open-source, developed into Mozilla
2003	Mozilla Foundation

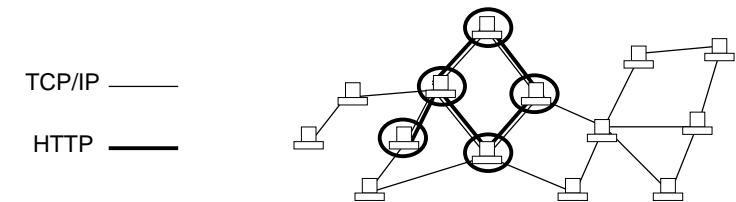**http://en.wikipedia.org/wiki/Web_browser**

Technical Details

HTTP
- Hypertext Transfer Protocol
- Language used to transfer Hypertext documents
- Client-Server Model

WWW
- Global body of information available using HTTP

TCP/IP ———

HTTP ▬▬▬

Cyberspace Addresses

Uniform Resource Locators (URL)
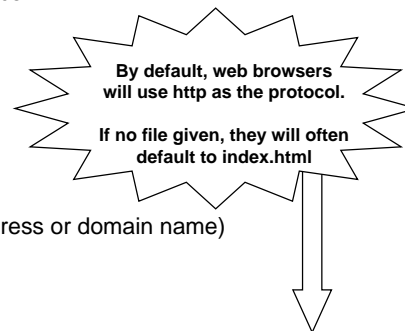- Address used for any web resource

Protocol
- Name of the protocol used
- ftp://   http://

Domain
- Name of a host computer (IP address or domain name)
- www.cs.auckland.ac.nz

File/ Resource
- Path of the file
- /andrew-l/index.html

> **By default, web browsers will use http as the protocol.**
>
> **If no file given, they will often default to index.html**

ihug.co.nz	**should be**
http://ihug.co.nz/index.html	

Terms

Web Site
- A collection of Web pages related to a single topic or theme. Normally designed and maintained by a single individual or organisation

Web Page
- A hypermedia document designed for the WWW

Web Browser
- Software used to access information on the World Wide Web
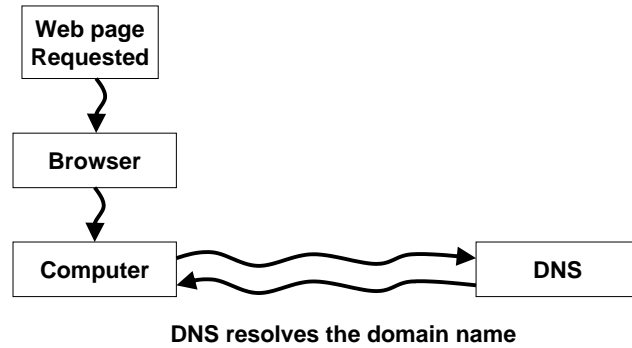- Sends requests to a web server
- Client

Web Server
- Software that makes local files available through the web
- Fulfils requests from a web browser
- Server

Accessing a web page (1)
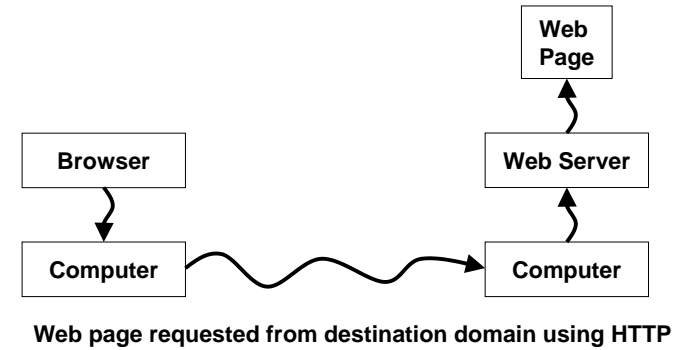
Client (Web Browser) runs on the local machine
- User requests a web page
- Client contacts the DNS to find the IP address

DNS resolves the domain name

Accessing a web page (2)

Web server runs on the destination machine
- Request sent to destination domain
- Web server accepts the request and finds the web page

Web page requested from destination domain using HTTP

Accessing a web page (3)

Web page is sent from the server to the client
- Client (web browser) displays the page

Web page sent from server to client using HTTP
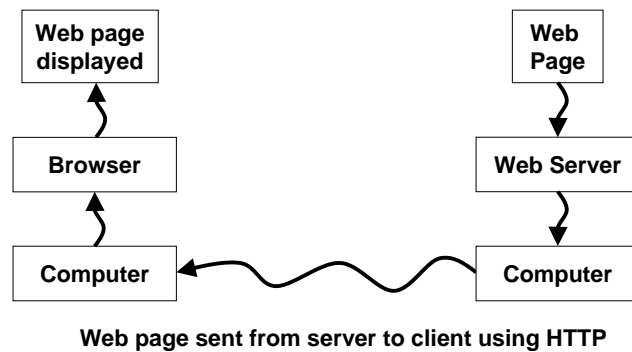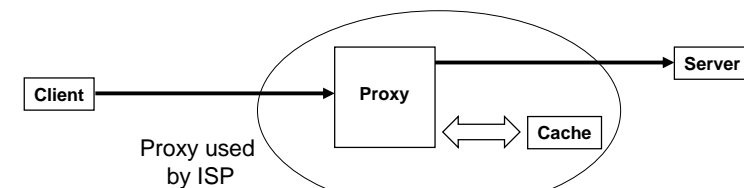
More Terms

Proxy
- A computer which sits between the client and server, intercepts and processes requests

Cache
- Store of information for quick access
- (e.g. caching may be used by proxy servers to speed web use)

Firewall
- Prevents unauthorised access to or from a private network

Proxy used by ISP

Logging web page access

Client keeps log
- History in web browser

Operating System keeps log
- Requests are logged by Windows on local machine

ISP keeps log
- Requests from "IP address" to "IP address" for "Page Name"
- Example, paradise.net has the logs available for users to check

Web server keeps log
- Gets requests from "IP address" for "Page Name"

Your viewing habits are being tracked!

http://en.wikipedia.org/wiki/Google_and_privacy_issues

Navigating

Finding information
- 45% users have problem finding new information
- 30% users have problems finding known information
- Web is very large, rapidly changing

Search Engines
- Automated
- Essential
- Our gateway to information

Search Engines

Companies
- Google
- Yahoo
- MSN Search

Automatically search every machine
- Archive the contents
- Index all the words
- Try to determine the relevance of the page

http://en.wikipedia.org/wiki/Search_engines

Can we trust the search engines?

Search Engines
- gateway to information
- pages are rated (how?)

Example: Veronica
- Camping search led to homosexuals
- Powerful groups lobbied
- Modified search engine (results filtered)

Publishers/ Advertisers
- Trick the search engines (repetition of words)
- Search engines tailor advertisements to searches
- Pay for higher rating?
- Who owns the search engines? How do they make their money?

http://en.wikipedia.org/wiki/Google_censorship

Recent Google News

Google sued for breach of copyright
- Claim fair use of images
- Appealing the decision

Google makes deal with Chinese Govt.
- Chinese Govt employs 30,000 people to censor the web
- Chinese version of Google is filtered

Brazilian Govt wants Google to help prevent online crime
- Google's social networking used to sell drugs

- Google and UMich

Searching

Searching Tips
- Learn how to use the advanced features of your search engine
- If the first page is not promising, choose different key words
- Use quote marks to search for phrases "Computer Science"

Finding useful sites
- Use specialist sites for specific searches
- Build a list of useful resources
- IMDB, New York Times, Ebay, BBC
- Trademe, NZ Herald, IRD, Woolworths, GOVT

Searching by URL
- Make a good guess
- Use knowledge of existing structure

Problems with the structure

Broken Links
- Pages which have been moved
- Referring page is not altered

No inherent security/ tracking/ accounting system
- Difficult to have layers of security
- Forces publishers to rely on advertising revenue

No inherent information indexing
- Much of the information is not accessed by search engines
- Information created on-the-fly from databases
- Information in other formats (postscript, pdf, archived)

Online innovations

Skype
- Voice over IP
- Talk to anyone in the world for free / cheap

Peer to Peer networks
- BitTorrent
- Swarming downloads

Google
- Acquired online word processor
- Rumours of alternate operating system
- Challenge Microsoft

Free Books
- http://digital.library.upenn.edu/books/
- http://books.google.com

Web-agents

Computer programs that operate on your behalf
- Tracks all your browsing habits
- Makes suggestions based on what you have read

Google Reader
- http://labs.google.com/

TiVo Suggestions
- Similar principle with television viewing

Future Directions

The only media of the future

The Internet is changing extremely rapidly
- 12 years since first appearance
- Too fast for legislation to keep up
- Too fast to predict the future

Coming soon
- Wireless Connections
- Wearable PC's
- Integrated Media (Interactive T.V.)
- Household Appliance connections
- Crime