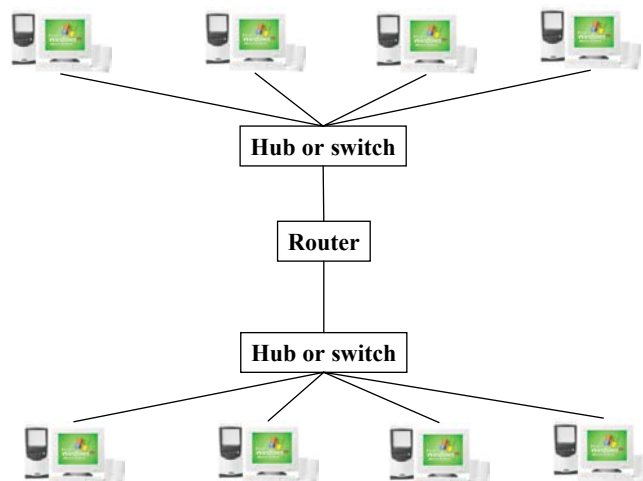


The Internet

**You will learn how the Internet works
and some ways of searching for
information on it.**

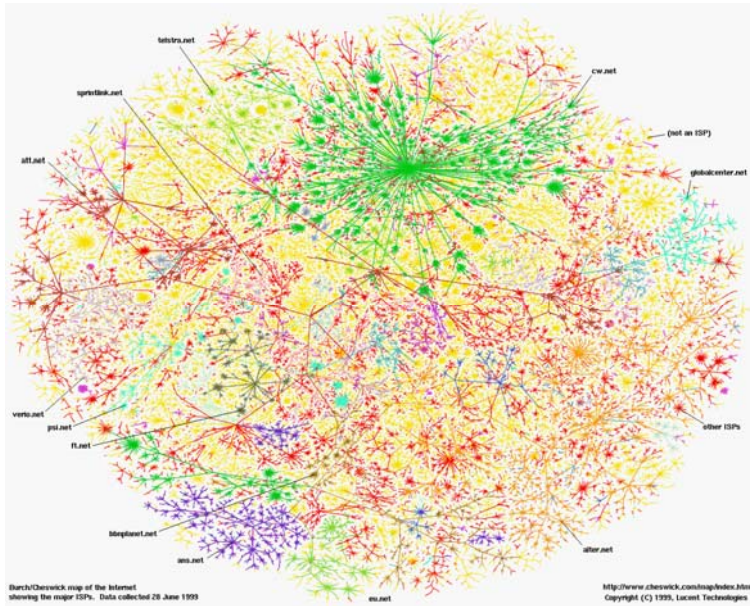
James Tam

A Network: Computers Which Are Connected



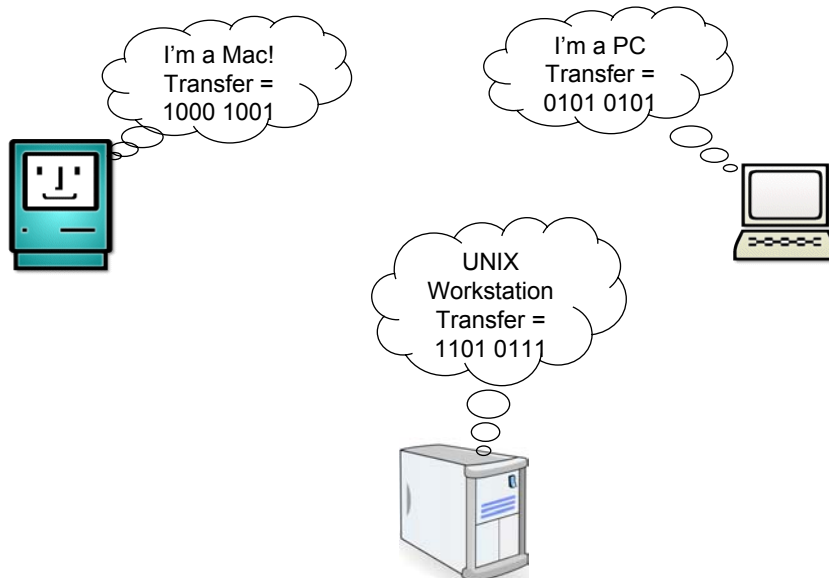
James Tam

The Internet: A Gigantic Network of Networks



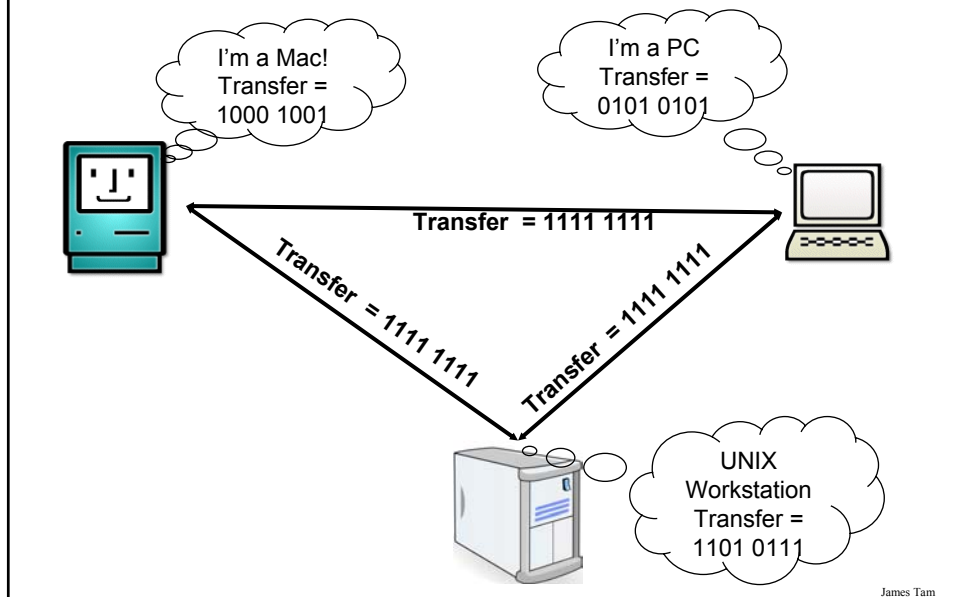
James Tam

Computers: Each Architecture Is Different



James Tam

Protocols: The Common Language Of The Internet

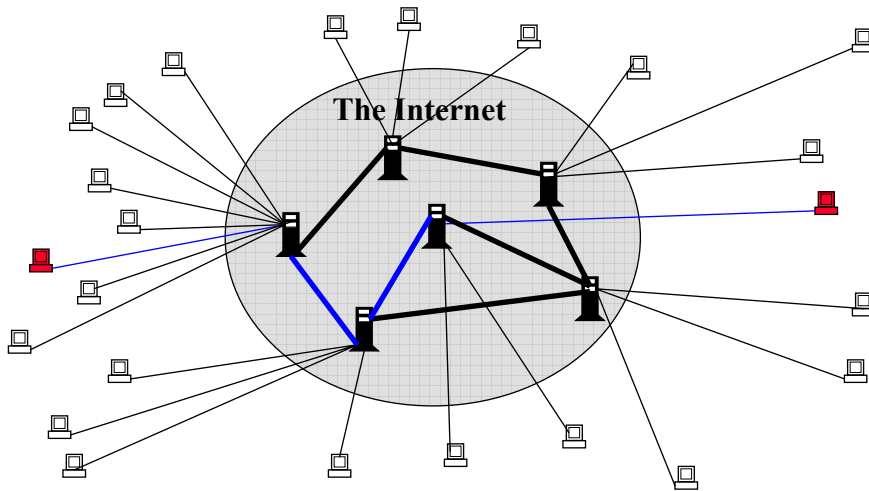


Protocols: The Common Language Of The Internet

(2)

- TCP/IP (Transmission control protocol / Internet protocol) is used for the transfer of information on the Internet.
 - TCP: is involved with disassembling/assembling the information being sent.
 - IP: is involved with ensuring the that information reaches it's correct destination.
- Other high-level protocols (easier to use) were developed for specialized tasks
- These protocols still employ TCP/IP but to complete the various tasks the higher-level protocols are the ones employed:
 - Http (Hypertext transfer protocol)
 - SMTP (Simple mail transfer protocol)
 - Telnet
 - FTP (File transfer protocol)

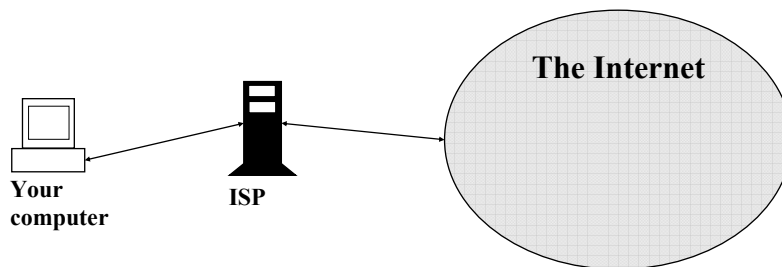
Speed Of Your Internet Connection



James Tam

Connecting Your Computer To The Internet

- Requires you to sign up with an Internet Service Provider (ISP)



- Types of Internet connections
 1. Old phone dialup connections (narrowband)
 2. Faster phone line connections (broadband)
 3. Cable connections (broadband)
 4. Satellite connections (broadband)

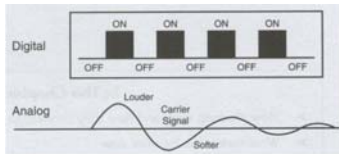
James Tam

1. Phone Dialup Connections

- Having your computer call another computer to connect to the Internet:



- Problem: Phone lines and computers don't easily mix

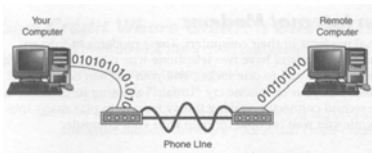


From the The Complete Idiot's Guide to High Speed Internet Connections by Soper

James Tam

1. Phone Dialup Connections (2)

- Having your computer call another computer to connect to the Internet
- Requires a modem (modulator/demodulator)



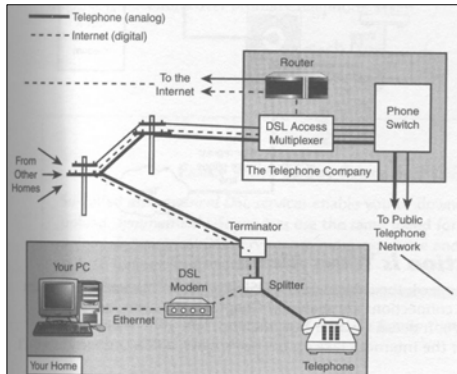
From the The Complete Idiot's Guide to High Speed Internet Connections by Soper

- Important characteristics:
 - Least expensive method for getting an Internet connection (free for you, for now): <http://www.ucalgary.ca/it/node/426>)
 - Very slow (300 – 56,000 bps/56Kbps)
 - Ties up a phone line
 - Not usable for many situations / not a continuous connection
 - Recommended usage: low bandwidth applications (text-only)

James Tam

2. Faster Phone Line Connections (DSL/ADSL)

- DSL (Digital Subscriber Line) / ADSL (Asymmetrical DSL)
- Uses a single wire for voice and data (computer)
- The phone lines are still used to transmit data but a different signal is used from standard dialup (faster)

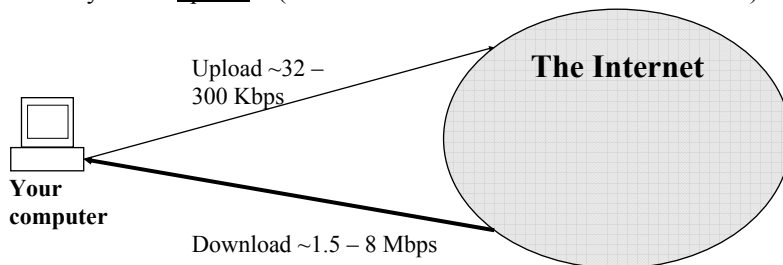


From the The Complete Idiot's Guide to High Speed Internet Connections by Soper

James Tam

2. Faster Phone Line Connections (DSL/ADSL)

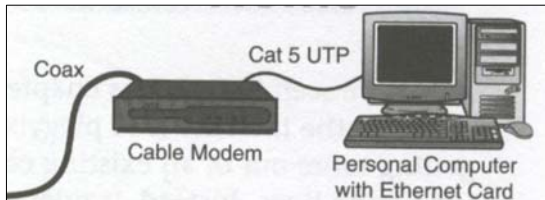
- Important characteristics:
 - Fast connection: ~ 100's Kbps – several Mbps
 - Doesn't use up a phone line
 - Allows for an always on (continuous) connection
- ADSL (Asymmetrical DSL)
 - You can download information (get information from the internet) faster than you can upload it (send information somewhere to the internet)



James Tam

3. Cable Connections

- Uses a coaxial (cable TV) connection.



From the The Complete Idiot's Guide to High Speed Internet Connections by Soper

- The Internet connection occurs on the same coaxial cable used for television.
- Important characteristics
 - Download: ~several Megabits (~1 – 6+ Mbps) per second.
 - Upload: approximately several hundred Kilobits (200 – 300 Kbps).
 - Allows for an always on (continuous) connection

James Tam

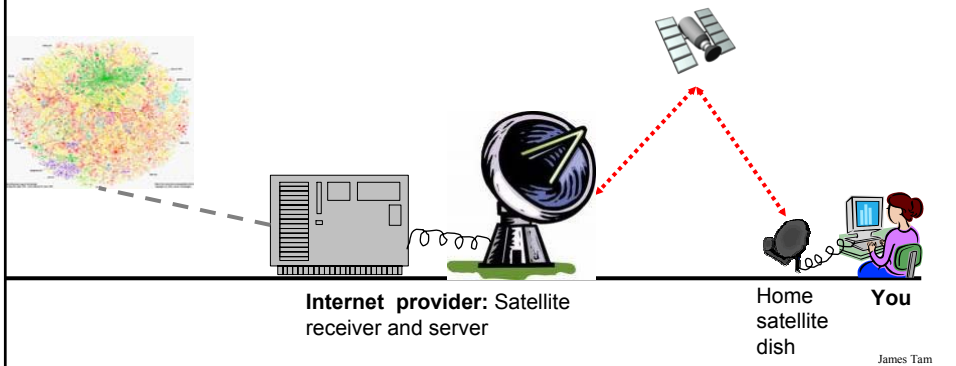
ADSL/DSL Vs. Cable

- | | |
|--|--|
| <ul style="list-style-type: none">• ADSL/DSL<ul style="list-style-type: none">- Typical transfer rates are slower- Internet connection is not shared- ADSL/DSL Internet service is highly restricted- (Calgary): The service is targeted more for 'experienced' computer users | <ul style="list-style-type: none">• Cable<ul style="list-style-type: none">- Typical transfer rates are faster- Internet connection is among shared subscribers- Cable Internet service is widely available- (Calgary): The service is targeted to the general population. |
|--|--|

James Tam

4. Satellite connections

- Typically if a broadband connection is desired then a standard wired ADSL/DSL or cable connection should be chosen.
- Some remote locations don't allow for either service (no cable wires or DSL/ADSL connections only standard dial-up).
 - In this case the only broadband connection available is via satellite:

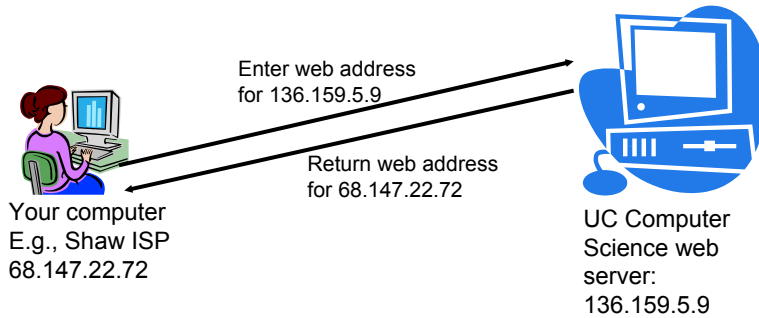


4. Satellite connections (2)

- Satellite connections are faster than dialup but slower than other broadband connections:
 - Upload ~ 100 Kbps
 - Download ~ 500 Kbps

IP (Internet Protocol) Addresses

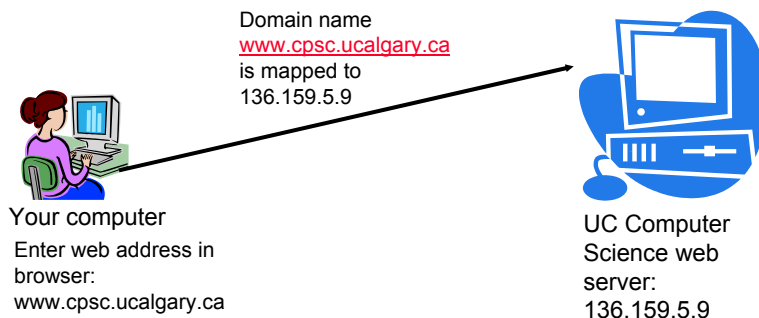
- Needed to properly route information on the Internet.
- Every computer connected to the Internet has an IP Address (IP for short):



James Tam

IP Addresses Are Mapped To Domain Names

- Because numeric addresses are hard to remember (and may change) domain names are typically used instead



James Tam

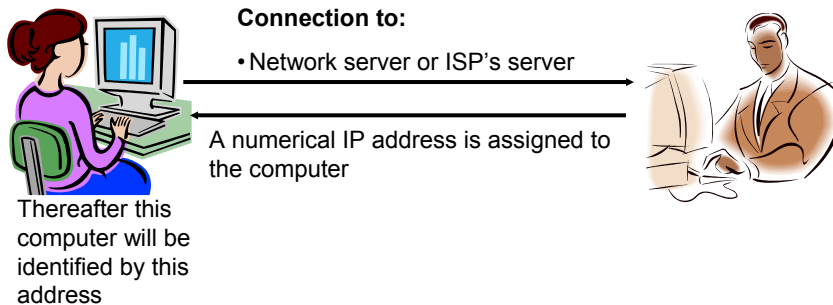
Assigning IP Addresses

- Different organizations (Registrars) provide Internet Addresses for Internet providers or websites:
 - VeriSign: dot-com (“.com”), dot-net (“.net”)
 - RegistryPro, NeuLevel and Public Interest Registry: dot-biz (“.biz”), dot-pro (“.pro”), dot-org (“.org”)
- When your computer gets an Internet connection it’s assigned an IP by your service provider which can be: dynamic or static.

James Tam

Static Addressing

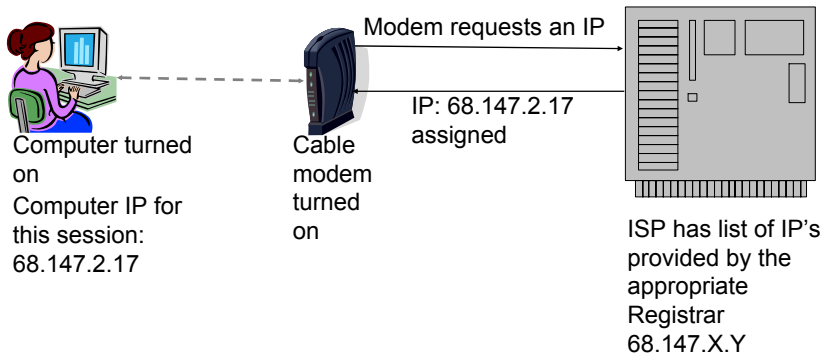
- When a computer is first connected to a network or the Internet it receives an IP address:



James Tam

Dynamic Addressing

- Whenever a computer is connected to a network or the Internet it's assigned a IP address from the existing pool of free addresses:



- Consequently the address may or may not be the same as ones that have been previously assigned.

James Tam

Dynamic Addressing (2)

- **Advantages:**

- Moving machines around the network to a different (sub)net is trivial. When the machine is connected at it's location it will learn of it's new IP.
- Resource efficient (not every computer needs it's own IP so fewer IP's are needed).

- **Disadvantages:**

- Some Internet services (e.g., web, ftp) require a fixed IP address which won't work with dynamic addressing.

James Tam

Static Addressing

- **Advantages:**

- Other computers may reliably make connections to this computer using certain protocols (e.g., FTP).

- **Disadvantages:**

- Machines cannot be simply be physically moved around the network to another (sub)net but must be electronically reconfigured.
- It's less resource-efficient (every computer must have it's own IP).
- It's less secure (the computer is "an unmoving target" because once a malicious program has found it, returning to that computer is trivial).

James Tam

URL's (Uniform Resource Locator)

- Provide a standardize way of describing information.
- Parts of a url (each part is enclosed by angled brackets):
<protocol> <machine where document is located> <path to the document>

e.g., <http://www.gamespot.com>

- [http](http://www.gamespot.com) indicates a hypertext document (a web page)
- www.gamespot.com domain on the world wide web (web server)

e.g., <ftp://ftp.ucalgary.ca>

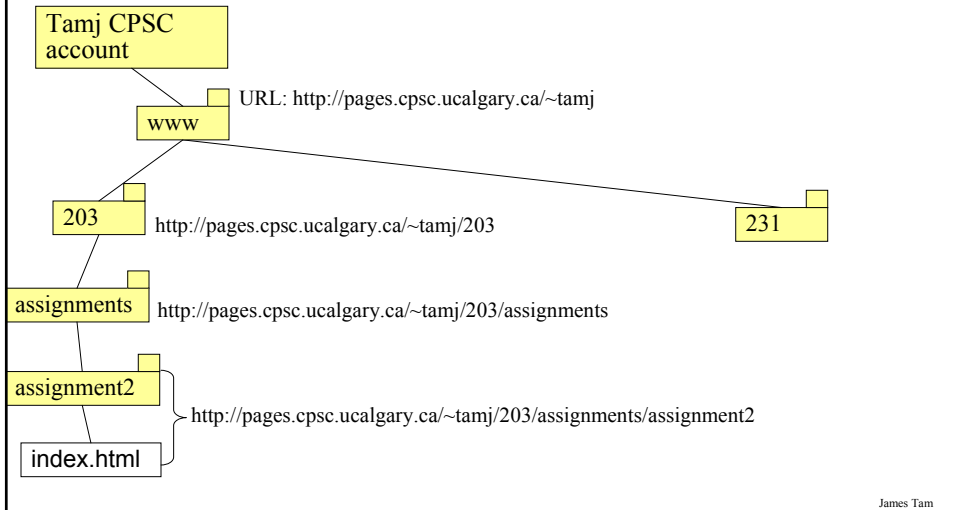
- [ftp](ftp://ftp.ucalgary.ca) indicates that the file transfer protocol is to be used
- [ftp.ucalgary.ca](ftp://ftp.ucalgary.ca) domain of the server (ftp server)

James Tam

URL's (Uniform Resource Locator): 2

- Can describe path to documents:

<http://pages.cpsc.ucalgary.ca/~tamj/203/assignments/assignment2/index.html>

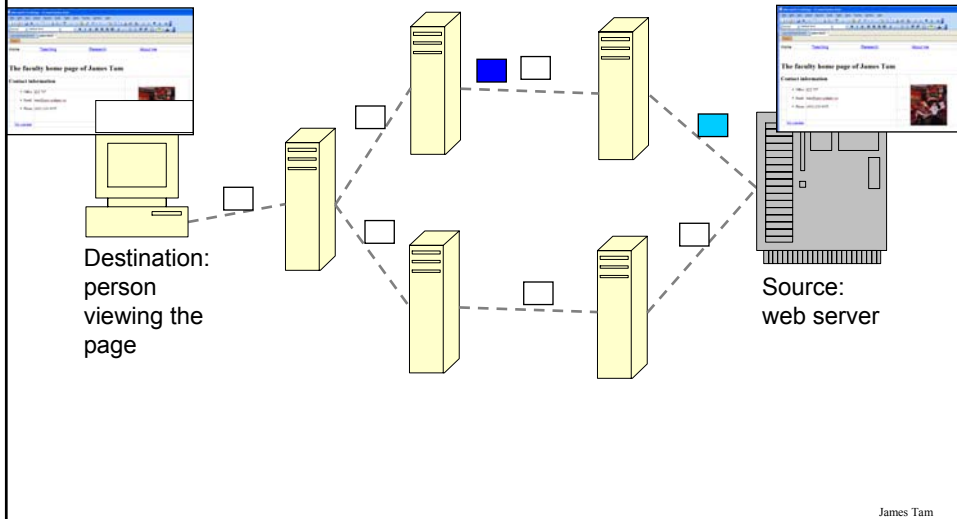


Transfer Of Information Via Packets

- The Internet is not an anonymous network!
- Information is transmitted via packets.
- In order to travel from their source to their destination each packet contains the following information:
 - **Header**
 - Source IP
 - Destination IP
 - Length
 - Number
 - Sequence
 - **Payload**
 - **Footer / Trailer**

Transfer Of Information Via Packets

- The route taken can vary from packet-to-packet

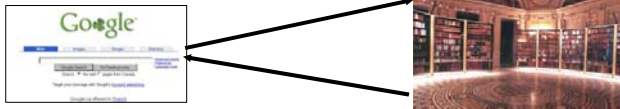


Intranets

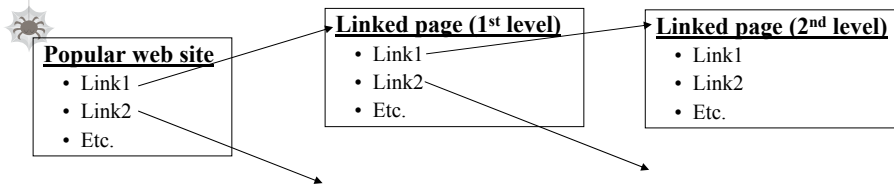
- A network that is a mini version of the Internet.
- Provides the features of the Internet (e.g., web pages) but is self enclosed:
 - Cannot be accessed from the outside.
 - Uses the protocols of the Internet.
- Extranet: the part of a company's or organization's intranet that is extended to non-employees or non-members.

How Do Search Engines Work

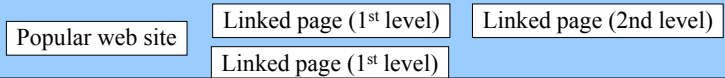
- Using a search engine involves searching the library of pages built up by that engine.



- The library is built up by the search engine's spiders (web)



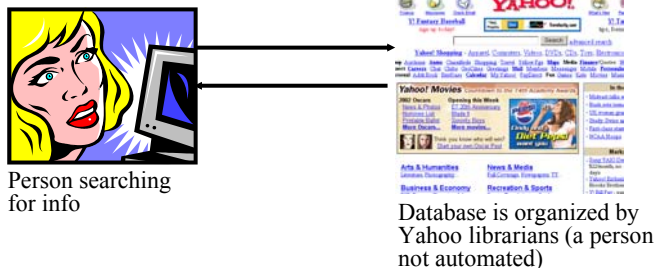
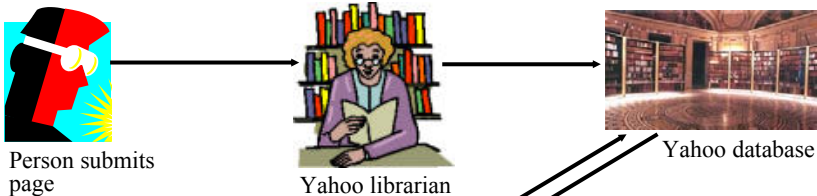
Search Engine Database



James Tam

Some Search Engines Use People Instead Of Technology

- Rather than using search programs to build the database some search engines organize results based on human scrutiny.



James Tam

Search Results Are Ranked

The screenshot shows a Google search results page for the query "James Tam". The search bar at the top contains "James Tam" and the search button is labeled "Search". Below the search bar, the text "Results 1 - 10 of about 16,600 for 'James Tam'. (0.06 seconds)" is displayed. The first result is titled "The faculty home page of James Tam" and includes the URL "tam@cpsc.ucalgary.ca". Other results include "The Scripps Research Institute - Faculty", "A little bit about James Tam!", "James Tam", "LinkedIn James Tam", "DBLP James Tam", "DBLP: Wa James Tam", and "W A James Tam on Find Articles". Each result includes a brief description and a "Cached" link.

James Tam

Making A Site More Noticeable

- Search database built via search spiders (e.g., Google):
 - Add relevant keywords to your page.
 - The frequency of keywords may play a role.
 - Engaging in spamdexing¹ may result in a page being put at the bottom of the list
- Search databases built via human researchers (e.g., Yahoo):
 - Make sure that your site is examined by the people who build the

The screenshot shows the Yahoo! "Suggest a Site" page. At the top, there is a "Sign In" button and a "New User? Sign Up" link. Below this, the text "Yahoo! Help > Directory Help > Suggest a Site" is displayed, with "Suggest a Site" circled in red. The main heading is "How do I add my web site to the Yahoo! Directory?". The text below explains that in order to add a web site to the Yahoo! Directory, the user must have their site added to the Directory. Two steps are listed: "Step 1: Check to see if your site is already in the Yahoo" and "Step 2: Find the appropriate category in the Yahoo! Dir".

¹ Spamdexing refers to the process of trying to artificially increase the ranking of a web page.

James Tam

Advanced Search Techniques (Google)

- Stopwords/Stop words and quotes
- Searching for synonyms
- Wild card searches
- Searching Ranges
- Site searches

James Tam

Stop Words

- Ignored by search engines:
 - Common words
 - Reserved words
- The search engine can be forced to include the stop words:
 - E.g., Use quotes (search results must contain whatever is between the quotes) or the 'plus' operator.

James Tam

Searching For Synonyms

- Some concepts can be represented using different words.
- The '~' operator includes synonyms in the search.

James Tam

Using The Wildcard In Searches



The screenshot shows a Google search interface. The search bar contains the text "arnold schw*" and is circled in red. The search results are displayed under the heading "Web". The first result is a PDF file titled "Page 1 2-19.pub" with a description: "26 Arnold Schwarzenegger, R, governor of California Tue., Feb. 27, Janet Napolitano, ... Arnold Schwarzenegger, R, will speak at a Club ...". The second result is a PDF file titled "BUSINESS" with a description: "Arnold Schwarzenegger signed a 'compact' providing for modest increases in the university budget. in exchange for commitments to make UC ...". The third result is a PDF file titled "1 2 SUPERIOR COURT OF CALIFORNIA COUNTY OF SACRAMENTO 10 11 ..." with a description: "ARNOLD SCHWARZENEGGER, Governor, in his individual capacity, GOVERNOR SCHWARZENEGGER'S CALIFORNIA RECOVERY TEAM, a ...".

James Tam

Searching Ranges

- Can be used when searching numerical values within a certain min – max range
- Range operator .. (multiple dots)

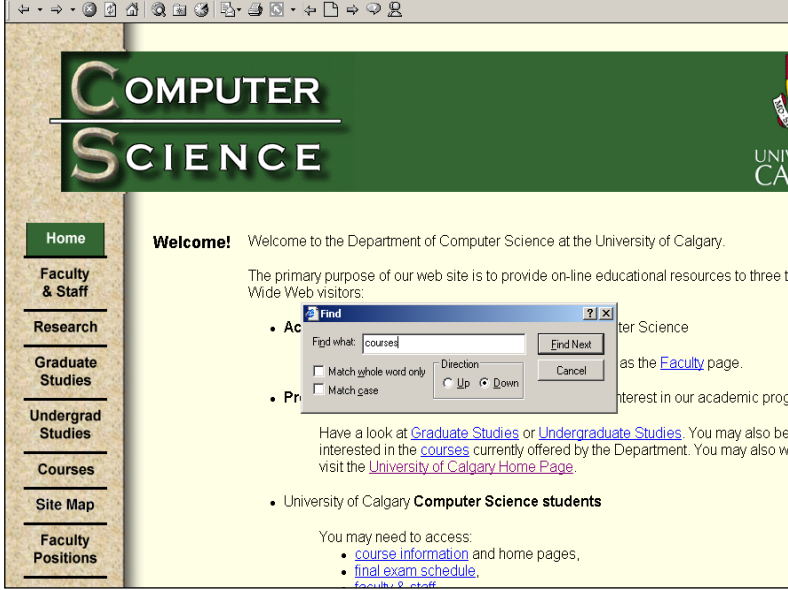
James Tam

Site Searches

- Useful when a webpage is large and/or not well organized:
 - Searching the currently loaded page
 - Searching the entire site

James Tam

Searching The Currently Loaded Page



The screenshot shows a web browser window displaying the University of Calgary Computer Science department page. A search dialog box is open over the page content. The dialog box has a title bar that says "Find" and contains the following elements:

- Find what:
- Match whole word only:
- Match case:
- Direction: Up Down
- Buttons: Find Next, Cancel

The background page content includes a navigation menu on the left with items like Home, Faculty & Staff, Research, Graduate Studies, Undergrad Studies, Courses, Site Map, and Faculty Positions. The main content area has a "Welcome!" message and a list of links for "Ac" and "Pr".

James Tam

Searching An Entire Site

- Involves searching an entire site (and not just the one page from the site that is currently loaded into the web browser).
- One of the 'advanced' search options can be employed.
- Alternatively a shortcut can be used in the main input field.

Meta-Search Engines

- Searches the databases of multiple search engines automatically.
- Examples:
 - www.metacrawler.com
 - www.dogpile.com
 - www.profusion.com
 - www.search.com
 - www.mamma.com
- Drawbacks:
 - Searches occur in the simplest form
 - Timeouts
 - Number of results returned

James Tam

You Should Now Know

- What is the Internet
- How protocols allows the Internet to function, when are the different protocols are used
- What are common ways of making an Internet connection, how do they work and what are their strengths and weaknesses
- How IP addresses are used to direct information on the Internet
- What is static vs. dynamic addressing and the strengths and weaknesses of each approach
- What is a url and what information is contained in a url
- How information is transmitted on the Internet via packets
- What is an Intranet

James Tam

You Should Now Know (2)

- How do search engines gather information
- Some ways of making a website more prominent to a search engine
- What is a natural language search, what is a good search engine to use when employing a natural language search
- Some advanced search techniques using Google:
 - What is a stop word and how to include them in searches
 - How to efficiently search for synonyms
 - How to employ a wildcard search
 - Searching ranges
 - Searching a specific site
- What is a meta-search and what are their weaknesses