

Announcements

- Remember: you are responsible for verifying what you submit through the submit server. You can download the zip file you submitted.
- Any questions about P3?

Searching the Web

- Many indices on the Web
- When visiting a server think of:
 - Who is providing the service
 - Whether they have some bias in the resources they catalog
- Starting points
 - General purpose search sites

www.google.com

www.altavista.com

Searching the Web

- Structured (and/or) Registration sites
www.lycos.com
www.yahoo.com
- Special purpose sites
www.switchboard.com – Yellow and White Pages
www.whowhere.com -
www.classmates.com -
www.findsounds.com/types.html

Searching the Web

- Different sites can crossover in purpose, services, abilities
- Major web sites are purchasing and/or integrating other sites in their domain. For example:
 - AOL purchase WebCrawler (www.webcrawler.com)
 - WhoWhere site became part of Lycos
- Big advantage of online indices: up-to-date information most of the time (unlike printed information)

Why searching is not trivial?

- Who do you trust?
 - Who owes the site?
 - Who created the page?
 - How to determine whether they are a trusted resource?
- How do you develop and refine searches?
 - identify keywords
 - add keywords or filter restrictions
- How do you search “concepts”?
 - “that movie with the girl and the dog and the red shoes”
 - “that painting by the guy that liked painting ballerinas”
 - “that song that starts dah-dah-dah-dum”

How to remember “Good” Sites

- Most web browsers provide a bookmarks
- Bookmarks are lists of URLs managed usually by the browser
- Problem with Bookmarks
 - Can be local you your system (e.g., bookmarks on your wam account stay with your wam account)
 - Bookmark lists may not move across web browsers
 - Firefox browser provides facility for bookmark
 - Firefox provides a “tab” feature for bookmark
 - Firefox allows you to bookmark your search engines
- On your Unix accounts adding a bookmark via Lynx will allow you to access the URL from any place that provides you with access to your account.
- You can use systems similar to google personalized page

Remote Access to Personal Preferences

- Several browsers are designed to access preferences (e.g. bookmarks and other customizable settings) from any Internet machine
- Netscape 4.5 is the first Netscape version to have this feature
- A technique used by many users is to create a web page with bookmarks
 - Advantage – bookmarks are available from any Internet machine
 - Disadvantage – bookmarking process is complex (web page modification)
- Google's personalized page allows you to create a bookmark set
- Windows Live Favorites (Beta) – Allows you to access your Favorites sites anywhere

Search Engines

- Popular engines
 - www.altavista.com (notice the category structure)
 - www.yahoo.com
 - www.metacrawler.com
 - www.ask.com (AskJeeves)
 - www.google.com
- <http://www.ipl.org/div/websearching/> - provides a good review of web search engines.
- Different syntax for search features
 - E.g. - www.altavista.com/help/search/syntax
- Common basic operation “and”, “or”, “not”, wildcards (*)
- In some search engines any word capitalization implies a case sensitive search
- Specifying details of a search is very important
- Ixquick Metasearch (<http://www.ixquick.com/>) - Relies on other search engines, translating your request so each search engine can understand it.
- Always look at the advance section of a search engine

Search Engines Tips

- Narrow search criteria – nothing will be found
- Broad search criteria – too many hits
- Some search engines support ability to ask questions (search engine discards common words)
- Example: Ask Jeeves (www.ask.com) or google

Google

- Access to more than 8 billion pages
- Features a built-in calculator
 - $\text{Sin}(30)$
 - $\text{Sqr}(25)$
 - pounds in a ton
 - “How many days in a century?”,
 - “How much is 30% of 2000?”
- Features several specialized search engines

Google Search Engines

- <http://labs.google.com/>
 - Provides prototypes (beta stage) for several products
 - <http://video.google.com/> (Search TV programs online)
 - <http://images.google.com/> (Search for Images)
 - Example: “Homer Simpson”
 - <http://groups-beta.google.com/> (Searchable groups/mailings lists)
 - <http://labs.google.com/sets> (Creates data set from examples)
 - Example: “Ringo Star” and “John Lennon”

Google Search

(From <http://www.google.com/help>)

- By default returns pages with all the specified terms (no need for “and”)
- Use OR to include either of two terms
 - E.g. vpn OR salsa
- Not case sensitive
- Common words (e.g., how and where) are automatically excluded
 - Add a “+” if you need the word
- “I’m Feeling Lucky” button – Returns most relevant site
 - E.g. University of Maryland
- For exact phrases use quotation marks (e.g., “Summer Party”)

Google Advanced Search

(From <http://www.google.com/help>)

- Advance operators (query words with special meaning)
 - **link** – Lists web pages with links to specified site
 - Example: link:www.cs.umd.edu
 - **info** – provides some information google keeps about a site
 - Example: info:www.cs.umd.edu
 - **define** – provides definitions
 - Example: define:atom
 - **allintitle** – all words must appear in the title
 - Example: allintitle:teaching science
 - **site** – search is limited to specified site
 - Example: Nelson Padua-Perez site: www.cs.umd.edu