Searching the WWW

Locating the right information on the WWW requires effort.

© 2004 Lawrence Snyder

Looking In the Right Place

The WWW is not the first place to look

• Go directly to a site --
  www.irss.gov
• Go to right sight --
  dictionary.cambridge.org
• Go to the library --
  www.lib.washington.edu
• Go for the kind of information you want --
  www.npr.org

Ask, “What site provides this information?”

Search Engines

No one controls what’s published on the WWW...it is totally decentralized.

To find out, search engines crawl Web

• Two parts
  • Crawler visits Web pages building an index of the content
  • Query processor checks user requests against the index, reports on known pages

Only a fraction of the Web’s content is crawled.

Google Advanced

Search for

Mona Lisa

Search Engine words are independent

• Words don’t have to occur together
• To be explicit about occurrences use Boolean queries and quotes

Logical Operators: AND, OR, NOT
  monet AND water AND lilies
  “van gogh” OR gauquain
  vermeer AND girl AND NOT pearl

Boolean Queries

Searching strategies...

• Limit by top level domains or format
• Find terms most specific to topic
• Look elsewhere for key words, e.g. bio
• Use exact phrase only when universal
• If too many hits, requery
• “Search within results” using “~”
• Once found, ask if site is best source

Queries
Truth on the Web

- Much Web information is wrong
- Using the Web effectively means recognizing quality information
  - Information from reliable organizations is usually preferred — check out ownership
  - Look for accuracy, currency, …
  - Follow links to verify that the content supports the original page

Best: Locate independent verification

A Bogus Site

The Burmese Mountain Dog

www.mnsu.edu/akcj3/bmd.html

Blue Site Bogus Implication

www.dhmo.org

Project 1

Create a bogus (fictitious) Web page

To appreciate how easy it is to fake quality info you will build a bogus Web page

- Modify photograph, changing its meaning
- Write misleading text
- Add “authenticity” links, fake credentials …
- Your page should look as legitimate as possible, but contain false information

Project 1 Details on Web

You will …

- Write page in HTML
- Find/take a photo
- Modify w/ Photoshop
- Find “authenticity” links
- Milestones
  - Turn in January 23: Web page plans
  - Turn in January 30: Web page + photo

Topics
Mars Rover
Iraq Reconstruction
UW Athletics Win
Democratic Race
California Recall
Recent Int'l Story
Research

Here’s How It Works

Follow these steps:

- Create your page locally
- When finished, publish it on the server
- Do not touch after the deadline (11:30PM)
- Print off a copy and turn it in in class
- Leave your page unchanged until grading is complete

We will check the timestamps