



## 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.iss.org](http://www.iss.org)
- Guessing a site's URL is often very easy, making it a fast way to find information
- Go to right sight -- [dictionary.cambridge.org](http://dictionary.cambridge.org)
- Go to the library -- [www.lib.washington.edu](http://www.lib.washington.edu)
- Go for the kind of information you want -- [www.nlm.nih.gov](http://www.nlm.nih.gov)

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

Google™ Advanced Search [Advanced Search Tips](#) | [All About Google](#)

Find results with all of the words with the exact phrase with any of the words without the words 10 results Google Search

Language Return pages written in any language

File Format Only return results of the file format any format

Date Return web pages updated in the anytime

Occurrences Return results where my terms occur anywhere in the page

Domains Only return results from the site or domain e.g. google.com, .org More info

SafeSearch No filtering Filter using SafeSearch



## Boolean Queries

Search Engine words are independent

Search for

- \* 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 gauguin
  - vermeer AND girl AND NOT pearl



## 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, query
- \* "Search within results" using "--"
- \* Once found, ask if site is best source



## 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 is a medium sized, muscular dog originally bred in Burma (Myanmar) to guard Buddhist temples. It was bred to guard the temples, and keep the temples free of rodents and beggars. It is also known as the Burmese Temple Dog. In 1954, a group of Burmese Ocean Lords set up a standard for the Burmese Mountain Dog which has remained virtually unchanged ever since. The Burmese Mountain Dog Club of America was established in 1985 to foster the breed in the United States and the world.

Do you want to own a Burmese Mountain Dog?

- The Burmese Mountain Dog is a breed of dog able to guard, herd small game, and protect property.
- The Burmese Mountain Dogs are remarkably clean dogs. They are easy to keep as they are rarely noisy or

[lme.mnsu.edu/akcj3/bmd.html](http://lme.mnsu.edu/akcj3/bmd.html)



## True Site, Bogus Implication

**DHMO.org**  
Dihydrogen Monoxide Research Division

**WELCOME**

Welcome to the web site for the Dihydrogen Monoxide Research Division (DHMO), currently located in Newark, Delaware. The controversy surrounding dihydrogen monoxide has never been more widely debated, and the goal of this site is to provide an unbiased data clearinghouse and a forum for public discussion.

Explore our many **Special Reports**, including the **DHMO FAQ**, a definitive primer on the subject, plus reports on the **environment**, **cancer**, **current research**, and an **index**.

[www.dhmo.org](http://www.dhmo.org)



## Project 1

### Create a bogus (fictitious) Web page

To appreciate how easy it is 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 Woes
- 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