



# Searching the WWW

*Locating the right information  
on the WWW requires effort*



# Looking In the Right Place

Google is not the first place to look

- Go directly to a site -- [www.irs.gov](http://www.irs.gov)

Guessing a site's URL is often very easy, making it a fast way to find information

- Go to right site -- [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.npr.org](http://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 (stored in a database)
- 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**<sup>TM</sup> [Advanced Search Tips](#) | [All About Google](#)

**Advanced Search**

**Find results**

with **all** of the words  10 results

with the **exact phrase**

with **any** of the words

**without** the words

**Language** Return pages written in

**File Format**  return results of the file format

**Date** Return web pages updated in the

**Occurrences** Return results where my terms occur

**Domains**  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



## Demonstration

- Google Images
  - \* 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, re-query
- \* "Search within results" using "-"



## Queries

- Once found, ask if site is best source
  - \* How authoritative is it?
  - \* Can you believe it?
  - \* How crucial is it that the information be true?
    - Cancer cure for Gramma
    - Hikes around Seattle





## 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



### Burmese Mountain Dog Guarding

#### [Gawdawpalin Temple](#)

---

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 Opium 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.

So you want to own a **Burmese Mountain Dog**?

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

<http://www.burmesemountaindog.org/>



# A Bogus Site

[Attention Students: Try our MLK Pop Quiz](#)

Martin Luther King, Jr.

## A True Historical Examination

That night King retired to his room at the Willard Hotel. There FBI bugs reportedly picked up 14 hours of party chatter, the clinking of glasses and the sounds of illicit sex - including King's cries of "I'm f--ing for God" and "I'm not a Negro tonight!"

Newsweek Magazine  
January 19, 1998



## Historical Writings

Essays, Sermons, Speeches & More

## Truth About King

Who He Fought and Fought For

## Death of the Dream

The Day King Was Shot

## The King Holiday

Bring the Dream to Life

## Civil Rights Library

History of People and Events

## Jews & Civil Rights

Who Led The Civil Rights Movement

## Suggested Books

<http://www.MartinLutherKing.org>



# True Site, Bogus Implication

The screenshot shows the homepage of DHMO.org. At the top left is the logo for the United States Environmental Assessment Center (EAC), featuring a sun and leaves. The main header reads "DHMO.org" in large blue letters, with "Dihydrogen Monoxide Research Division" below it. To the right is a circular badge that says "Support the cause! Visit the DHMO.org Store" with logos for VISA, PayPal, and MasterCard. The page is divided into three columns. The left column, titled "DHMO Special Reports", lists several links: "Dihydrogen Monoxide FAQ", "Environmental Impact of DHMO", "Dihydrogen Monoxide and Cancer", "DHMO Surveys & Research", "DHMO in the Dairy Industry", "DHMO Conspiracy", "Editorial: Truth about DHMO", "Fake Email SPAM Alert", and "Linking to DHMO.org". The middle column, titled "WELCOME", contains a paragraph: "Welcome to the web site for the Dihydrogen Monoxide Research Division (DMRD), 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." Below this is another paragraph: "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 insider." The right column, titled "DHMO Related Info:", lists several links: "Media Press coverage", "National Consumer Coalition Against DHMO", "Green Party, New Zealand", "Environmental Protection Agency", and "NIH National Toxicology Program". At the bottom of the screenshot is a blue banner with the URL "http://www.dhmo.org" in white text.

**EAC**  
United States  
Environmental  
Assessment  
Center

## DHMO.org

Dihydrogen Monoxide  
Research Division

Support the  
cause! Visit the  
DHMO.org  
Store

VISA PayPal MasterCard

### DHMO Special Reports

- [Dihydrogen Monoxide FAQ](#)
- [Environmental Impact of DHMO](#)
- [Dihydrogen Monoxide and Cancer](#)
- [DHMO Surveys & Research](#)
- [DHMO in the Dairy Industry](#)
- [DHMO Conspiracy](#)
- [Editorial: Truth about DHMO](#)
- [Fake Email SPAM Alert](#)
- [Linking to DHMO.org](#)

### WELCOME

Welcome to the web site for the Dihydrogen Monoxide Research Division (DMRD), 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 [insider](#).

DHMO Related Info:

- [Media Press coverage](#)
- [National Consumer Coalition Against DHMO](#)
- [Green Party, New Zealand](#)
- [Environmental Protection Agency](#)
- [NIH National Toxicology Program](#)

<http://www.dhmo.org>



## April Fool's Prank Site

“The US National Institutes of Health is to crack down on scientists ‘brain doping’ with performance-enhancing drugs such as Provigil and Ritalin, a press release declared last week. The release, brainchild of evolutionary biologist Jonathan Eisen of the University of California, Davis, turned out to be an April Fools’ prank. And the World Anti-Brain Doping Authority website that it linked to was likewise fake.”

—*Nature* magazine



# April Fool's Prank site

## World Anti-Brain Doping Authority (WABDA)



- » [Home Page](#)
- » [Mission](#)
- » [News](#)
- » [Contact Us](#)
- » [The WABDA Code](#)

### Welcome

WABDA is an independent foundation created through a collaboration between the National Institutes of Health (NIH) in the United States of America, the European Commission and the World Anti-Doping Authority (ADA). It was set up on January 10, 2008 to coordinate the fight against brain doping in Academia. Its current chairman is Richard Pound, the current head of WADA.

The agency works to help individual academic federations implement testing procedures in the fields of academic research. It also produces a list of prohibited substances that academics are not allowed to take and maintains the World Anti Brain-Doping Code.



Copyright 2008. WABDA - World Anti Brain-Doping Agency. All rights reserved.

[http://wabda.org/Home\\_Page.html](http://wabda.org/Home_Page.html)



## Unit I Project

### 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 ...



## Unit I Project

- \* Your page should look as legitimate as possible, but contain false information
- \* A site visitor should start out believing your site, but by the time they finish reading, they should realize that it's a hoax
- \* Forget subtlety!





## Turn In Steps

- Publish your page by uploading to the Web server
- copy your Project1A files into a Project1B folder
- Submit your Word or .txt with project URL in Catalyst Collect It
- Do not touch anything in your 1A folder after the deadline
  - \* So the TAs can grade it

We will check  
the timestamps