# **Research-quality Web Searching:**

**Google and Beyond**

## John Kupersmith

## jkupersm@library.berkeley.edu

##  A “Know Your Library” Workshop

## Teaching Library, University of California, Berkeley

## Spring 2009

# COURSE PAGE:

www.lib.berkeley.edu/find/types/websites.html

ONLINE TUTORIAL WITH MORE DETAILED INFORMATION:

# www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html

**Goals for this workshop**

## Search Google effectively and precisely

## Know when to use other search engines and web directories

## Evaluate what you find on the web

# **How Google works**

## BEFORE you search

## “Crawls” pages on the public web

## Copies text & images, builds database

## WHEN you search

## Automatically ranks pages in your results

### Word occurrence and location on page

### Popularity - a link to a page is a vote for it

### ~ 200 factors in all!

# **Searching Google**

## Think “full text” = be specific **war of 1812 economic causes** vs. **history**

## Use academic & professional terms**domestic architecture** vs. **houses****genome society**gets  *International Mammalian Genome Society*also try combinations with **association, research center, institute, directory, database**

## Specify exact phrases **“tom bates”** **“what you're looking for is already inside you”**

## Exclude or require a word **proliferation -nuclear bush legacy +environment**

# **Limit your search to …**

## Web page title**intitle:hybrid allintitle:hybrid mileage**

## Website or domain**site:whitehouse.gov “global warming”site:edu “global warming”**

## File type**filetype:ppt site:edu “global warming”**

## Definitions**define:pixeldefine:“due diligence”**

# **On the results page**

## Search box (use to modify your search)

## “Cache”“Related pages”“Translate this page”

# **Google’s other databases**

Images, Maps, News, Blogs, Books, Scholar, etc.

# **Why go beyond Google?**

## Search more of the web **Yahoo!** -- search.yahoo.com

## Get more options  **Exalead** -- www.exalead.com/search

## Take advantage of human selectivity

##  **Librarians’ Internet Index** -- www.lii.org **InfoMine** -- infomine.ucr.edu **Google Custom Search Engines (CSE)**

## How to find these: www.lib.berkeley.edu/find/types/websites.html

##  Scroll down to “New Approaches to Web Searching”

# **Critical Evaluation:Why Evaluate What You Find on the Web?**

## Anyone can put up a web page

## Many pages not updated

## No quality control

### most sites not “peer-reviewed”

#### less trustworthy than scholarly publications

# **Web Evaluation Techniques:Before you click to view the page...**

## Look at the URL - personal page or site ?  **~** or **%** or **users** or **members**

## Domain name appropriate for the content ?

### Restricted: **edu**, **gov**, **mil**, a few country codes (**ca**)

### Unrestricted: **com**, **org**, **net**, most country codes (**us**, **uk**)

## Published by an entity that makes sense ?

#### News from its source?

######  **www.nytimes.com**

#### Advice from valid agency?

######  **www.nih.gov/**

######  **www.nimh.nih.gov/**

# **Scan the perimeter of the page**

## Can you tell who wrote it ?

#### name of page author

#### organization, institution, agency you recognize

## Credentials for the subject matter ?

### Look for links to:

### **“About us” “Philosophy” “Background” “Biography”**

## Is it current enough ?

#### Look for **“last updated”** date

## **Examine the content**

## Textpossibly forged ?

#### why not a link to published version ?

* Sources

#### documented with links or notes ?

#### do the links work ?

## Evidence of bias

####  in text or sources ?

# **Do some detective work**

## Search the URL in **Alexa** -- www.alexa.com

### Click on “Site info for …”

### Who links to the site?

### Who owns the domain?

### What did the site look like in the past?  (use the “Wayback Machine” link)

## Which blogs link to it? What do they say?

### Try the URL in **Google Blog Search** -- blogsearch.google.com

## See what links are in Google’s “Similar pages”

## Look up the page author in Google

# **Does it all add up ?**

## Was the page put on the web to

#### inform ?

#### persuade ?

#### sell ?

#### as a parody or satire ?

## Is it appropriate for your purpose?

# **Try evaluating some sites**

## Search a controversial topic in Google

### **nuclear armageddon**

### **prions danger**

### **“stem cells” abortion**

## Scan the first two pages of results

## Visit one or two of these sites

### Use checklist (next page) to evaluate their quality and reliability

2/27/09

Google and Beyond Copyright © 2012 The Regents of the University of California is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](http://creativecommons.org/licenses/by-nc/3.0/). Permissions beyond the scope of this license may be available at  <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/contact.html>.