

# Google and Beyond: Research-Quality Web Searching

John Kupersmith

[jkupersm \[at\] library.berkeley.edu](mailto:jkupersm@library.berkeley.edu)

A “Know Your Library” Workshop  
Teaching Library, University of California, Berkeley  
Spring 2009

## COURSE PAGES:

<http://www.lib.berkeley.edu/find/types/websites.html>

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>



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>.

# Goals

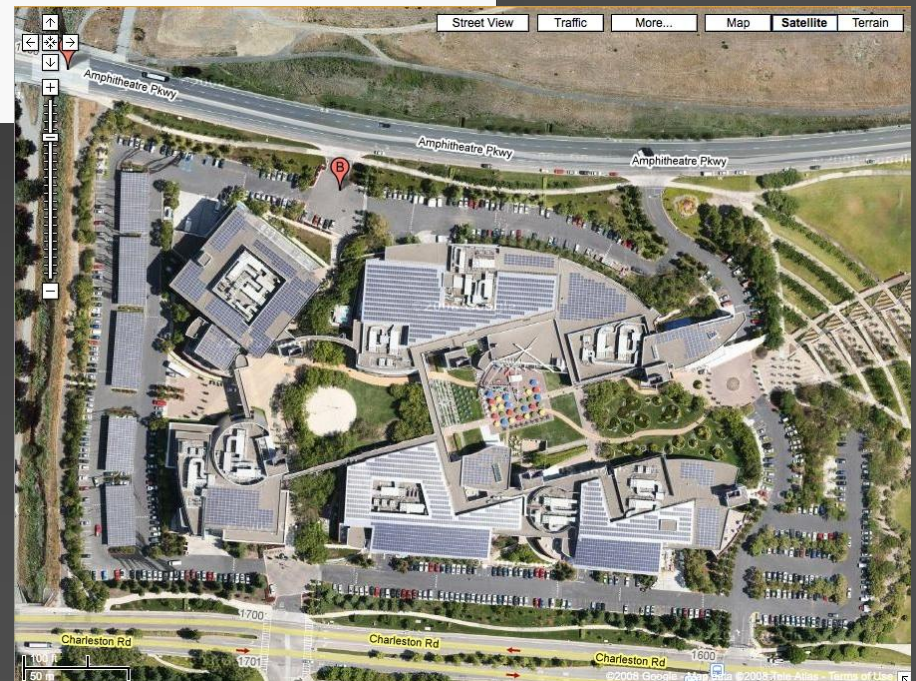
- Search Google effectively and precisely
- Know when to use other search engines and web directories
- Evaluate what you find on the web



[Advanced Search](#)  
[Preferences](#)  
[Language Tools](#)

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2008 Google



# 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**

# Searching Google

- 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"`

# Limit your search to ...

- File type

**filetype:ppt site:edu “global warming”**

- Definitions

**define:pixel**

**define:“due diligence”**



# On the results page

- Search box (use to modify)
- “Cache”
- “Related pages”
- “Translate this page”

# Google's other databases

The image shows a screenshot of the Google homepage from 2008. At the top, there are navigation links: [Web](#), [Images](#), [Maps](#), [News](#), [Shopping](#), [Gmail](#), and [more](#) with a downward arrow. Red arrows point to [Images](#), [Maps](#), and [more](#). The [more](#) dropdown menu is open, listing: [Blogs](#), [Books](#), [Calendar](#), [Documents](#), [Finance](#), [Groups](#), [Photos](#), [Reader](#), [Scholar](#), [Video](#), and [YouTube](#). A red arrow points to [Scholar](#). A blue arrow points to the [more](#) dropdown menu. Below the search bar, there are links for [Advanced Search](#), [Preferences](#), and [Language Tools](#). At the bottom, there are links for [Advertising Programs](#), [Business Solutions](#), and [About Google](#). The copyright notice [©2008 Google](#) is at the very bottom.

# Why go beyond Google?

- Search more of the web

Yahoo! search.yahoo.com

- Get more options

Exalead www.exalead.com/search

# Why go beyond Google?

- Take advantage of human selectivity

[Librarians' Internet Index](#)

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

[InfoMine](#)

[infomine.ucr.edu](http://infomine.ucr.edu)

[Google Custom Search Engines \(CSE\)](#)

**How to find these:**

[www.lib.berkeley.edu/find/types/websites.html](http://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

# 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](http://www.nytimes.com)**
  - Advice from valid agency?  
**[www.nih.gov/](http://www.nih.gov/)**  
**[www.nimh.nih.gov/](http://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

## ■ Text

- possibly 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.com](https://www.alexa.com)
  - Click on “Site info for ...”
  - Who owns the domain?
  - Who links to the site?
  - What did the site look like in the past?  
(Wayback Machine)

# Do some detective work

- Which blogs link to it? What do they say?
  - Try the URL in [Google Blog Search](#)
- 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...

1. Search a controversial topic in Google such as
  - **nuclear armageddon**
  - **prions danger**
  - **“stem cells” abortion**
2. Scan the first two pages of results
3. Visit one or two sites
  - Use the checklist (next two slides) to evaluate their quality and reliability

# Web Page Evaluation Checklist



<b>Title and URL of page you are evaluating:</b>	
<b>Personal page or site?</b>	<input type="checkbox"/> ~ or %, or <i>users, members, or people</i>
<b>What type of domain is it?</b> Appropriate for the content?	<input type="checkbox"/> .com <input type="checkbox"/> .org/net <input type="checkbox"/> .edu <input type="checkbox"/> .gov/mil/us <input type="checkbox"/> non-US _____ <input type="checkbox"/> other:
<b>Published by entity that makes sense?</b> Does it match the name of the site?	Publisher or Domain Name entity:
<b>Who wrote the page?</b>	<input type="checkbox"/> E-mail <input type="checkbox"/> Name:
<b>Credentials on this subject?</b> (Truncate back the URL if there are no useful links.)	Evidence?
<b>Dated?</b>	Date _____ Current enough for your purpose?

# Web Page Evaluation Checklist

<b>Sources documented with links or notes?</b> Can you verify that quoted text was not altered or forged?	
<b>Links to more resources?</b> Do they work?	
<b>Evidence of bias?</b>	
<b>Search URL in alexa.com, click on "Site info for ..."</b>	Who owns the domain? Who links to the site?
<b>Is the page rated well in a directory?</b> www.lii.org or infomine.ucr.edu or about.com	
<b>Which blogs link to it? What do they say?</b> blogsearch.google.com	
<b>Look up the author in Google</b>	
<b>Why was the page put on the Web?</b>	<input type="checkbox"/> Inform <input type="checkbox"/> Persuade <input type="checkbox"/> Sell <input type="checkbox"/> Satire or parody? Other:

## Bottom Line:

Is the web page appropriate for  
your purpose?