



Open Source Center

---

# Advanced Googling for Senior Executives



Open Source Academy  
*Internet Science Faculty*

---

September 2009

- Intros
- Basic & advanced search
  - Features, operators, and syntax
  - Google focused, but applies to other search engines.
- Hands-on class
- Explore, experiment, have fun!
- Questions? Comments?

Not an endorsement of Google

# Instant Gratification

- Google provides many useful features

- Definitions
  - define:ontology
  - define:mahdi
- Global time
  - time Tehran, Iran
- Global weather
  - weather Paris
- Calculator
  - 25% of 79.99
- Unit conversions
  - 1 foot 3 inches in meters
  - 100 USD in Pakistan rupee
- Area codes
  - area code 571
- Synonym search
  - ~car loan
- Spell checking
  - Misissippi

<http://www.google.com/help/cheatsheet.html>  
<http://www.googleguide.com/help/calculator.html>

- There are many search engines on the Internet
  - Examples? Capabilities?
- There are many sites that search engines can't search
  - What are some examples?
- The Internet is not the final answer; some things are just not there
  - Librarians can search databases you cannot!
  - Who's your LIBRARIAN?



# Now that we're hooked, let's step back and ask a few questions ...

- What is the Internet & what isn't it?
  - It is a standardized global system of interconnected computer networks that connects millions of people; not all networks connected
  - The Internet is not the final answer; some things just aren't there
- Are search engines always the best answer to a problem?
  - Specialized/vernacular web directories may be the fastest way to research some subject; librarians are often the best answer
- Can these techniques be used elsewhere?
  - Appliances, or ports of search engines to private networks, will sometimes lack features or behave differently
  - Other search engines may have slightly different syntax, check help pages

# Caveats!!!

- We cannot cover all aspects of search in one class
  - Internet and all search engines changing at Internet speed
  - Read the manual or help pages even though they're hard to find
- Experiment with more than one search engine
  - Why?
- Some things can't be explained
  - Dig deep or keep reading
  - Might find answer or maybe not or...
  - It may not matter!
- Techniques can be used with appliances, other engines
  - Appliances--ports of search engines--lack features or behave differently
  - Other search engines may have slightly different syntax, check help pages

# Set Google Preferences

- Go to Preferences
- Choose your Interface Language
  - On View menu, choose Encoding, (Unicode) UTF-8
- Search Language: set to any language
- SafeSearch Filtering: Do not filter my results
  - Depending on your cookie settings, this may be reset
- Number of Results: Display 100
- Results Windows: your choice; Click Save Preferences

more »



[Advanced Search](#)

[Preferences](#)

[Language Tools](#)

# Search Strategies

- You're getting ready to leave your house. You can't remember what you did with your keys. What do you do? What's the point?
  - Pick peculiarity, turn your question into a query
- For this seminar, we define a continuum of search strategies:
  - No thought required (NTR): just type in your terms
  - Basic: using basic Boolean operators, " "
  - Advanced: using Boolean operators and advanced operators/syntax

Search is often an iterative process!



- No thought required (NTR)

- credit report request form

- Often NTR produces too many hits: e.g. virus \_\_\_\_\_

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

- Basic

Results 1 - 100 of about 197,000,000 for virus

- AND (no need to type the AND operator)

- Space defaults to AND most search engines

- computer virus

- kim jong il

- OR (must be typed and capitalized)

- computer virus OR worm

- kim jong OR chong il

- Basic

- - (exclusion operator aka NOT)
- Always a space before and never a space after
  - virus –computer
  - bass –fish
  - osama –bin -laden
- " " (phrase search)
  - "computer virus" -"trojan horse"
  - "kim jong il" OR "kim chong il"
  - "ied stands for" explosive
- .. (number range)
  - "digital camera" "8..12" megapixels ..\$1000

- Basic: taking it further
  - \* ("fill in the blanks" operator; not a "wildcard")
    - "love or \* must constantly"
    - "senior \* service"
    - "pgt \* amplifier"
  - + (inclusion operator)
    - star wars +l
  - ~ (synonyms and related words)
    - matrix ~organization -math
- Peculiarity and patterns are profound: most powerful thing you can take away from this class
  - Peculiar: gadolinium
  - Pattern: en in a url

- Google ignores punctuation, except
  - Punctuation in popular terms: C++ C#
  - " - \$ : . \_
  - What was the single-dash NOT rule?
- Capitalization does not matter
  - Except for operators: site not SITE
- No spaces before or after colons



- Think about how to turn your natural language question into a query before hitting enter
- Think about how you would apply the advanced operators at work or at home
- We will not cover all advanced operators

# Select Advanced Operators

- We will limit our class to seven operators
  - site
  - inurl, intitle, intext
  - filetype
  - date
  - info

## Rocket Motors

**Rocket Motors** burn solid grain propellant to generate large amounts of thrust. Launch vehicle designers typically have to decide between using liquid rocket ...

[www.spaceandtech.com/spacedata/motors/motors.shtml](http://www.spaceandtech.com/spacedata/motors/motors.shtml) - 12k - [Cached](#) - [Similar pages](#)

- Title: **Rocket Motors**
- Snippet (text): **Rocket Motors** burn solid grain propellant to generate large amounts of thrust. Launch vehicle designers typically have to decide between using liquid rocket ...
- URL: [www.spaceandtech.com/spacedata/motors/motors.shtml](http://www.spaceandtech.com/spacedata/motors/motors.shtml)
- Size: 12k
- Cached - Similar Pages: Instructor will explain

# Common Errors

- RIGHT      `intitle:"politcal stability"`
- WRONG      `intitle:⌘political stability"`
- WRONG      `"intitle:political stability"`
- RIGHT      `"ammonium diuranate" -site:com`
- WRONG      `"ammonium diruanate -⌘site:⌘com"`

⌘ = space



- **site** searches within a specific website
  - Learn **site** well and other operators come easy
    - `iraq site:www.cnn.com` scan URL results
- **site** can be used to search top-level domains
  - Search & compare # of hits
    - `"brake pads"` \_\_\_\_\_
    - `"brake pads" -site:com` \_\_\_\_\_
    - `"brake pads" site:edu` \_\_\_\_\_

- **site** can also narrow search to a country domain
  - Jump start research on a topic and country:
    - "opposition party" **site:my**
  - How could you find a list of internet country codes?
  - Two examples of global reach
    - **your\_search\_terms site:xx**
      - Search the world from your desk by changing just two letters
    - **site:jp**, go to Advanced Search, select Arabic
      - What did we do? What did we find?
      - How else can you use this technique?

- Country code doesn't mean server is in country
  - site is registered to xx country domain
  - A Chinese site (cn) could be hosted in Canada
- **site:ws**
  - Why might some of these sites be registered in Western Samoa?
- Get specific indexed data from a web site
  - **site:kadena.af.mil** then click on Images
  - **850 OR 381 site:kcna.co.jp**

- More **site** examples
  - Build **site** queries from right to left
    - **site:apple**

---
    - What happened?
    - Compare results and explain
      - **site:apple.com**

---
      - **site:www.apple.com**

---



- This is a URL:

- protocol://computer.domain.name/pathname/filename.ext
  - For our purposes, site = computer domain name

- Thinking back to the missing keys example
  - Where in the URL does the **site** operator search?

[Rocket Motors](#)

**Rocket Motors** burn solid grain propellant to generate large amounts of thrust. Launch vehicle designers typically have to decide between using liquid **rocket** ...

[www.spaceandtech.com/spacedata/motors/motors.shtml](http://www.spaceandtech.com/spacedata/motors/motors.shtml) - 12k - [Cached](#) - [Similar pages](#)

- Background processing: site & URL, difference?

- Searches the whole URL (i.e. searches urls of web pages)
- inurl can be useful after you discover a pattern in preliminary search results
  - site:kcna.co.jp
    - Scan URL results for patterns
  - site:kcna.co.jp inurl:item inurl:2009

- inurl queries using *a priori* (i.e. reasoning/logic/rationalism methodology) info
  - inurl:meshawir
  - inurl:2009 site:state.gov
  - inurl:blog site:xx (choose country code)
  - inurl:english site:xinhuanet.com currency

- What's the difference between site & inurl?
  - site:www.state.gov \_\_\_\_\_
  - inurl:www.state.gov \_\_\_\_\_
  - **site** searches everything before the first /
  - **inurl** searches the whole URL including everything before and after the first /
  - Unusual to have a site refer back to itself  
[Countries, Regions](#)  
Page links to a variety of subjects related to countries and regions of the world.  
[www.myday.com.tw/guest/filter.php?http://www.state.gov/countries](#) - 27k - [Cached](#) - [Similar pages](#)
  - but not **www.state.gov/../../www.state.gov**



- The previous material highlights the awesome power of advanced Googling
- Patterns can be exploited to find information that Google has not yet indexed!
  - Bangladesh Chittagong Port Authority site
    - site:cpa.gov.bd inurl:berthing
    - site:cpa.gov.bd inurl:berthing inurl:yymmdd
    - www.cpa.gov.bd/berthing/yymmdd-berthing.html

- *intitle* restricts search to web page titles
  - `intitle:"inspector general" site:treas.gov`
  - Useful for specific topic or patterns in preliminary search
    - `avian influenza site:oie.int`
    - `avian influenza site:oie.int intitle:"disease information"`
    - `avian influenza site:oie.int intitle:"disease information" 2005..2007`
    - `site:cpa.gov.bd intitle:"index of"`

- **intext** searches for text in body
  - Good for finding e-mail addresses, zip codes, phone numbers, and other information that is URL-like; may help with defunct websites
    - `intext:ngsec.com`
    - `intext:opensource.gov`
    - `intext:"internal use only" site:gov`
    - `intext:gmail.com` OR `intext:hotmail.com site:sy`

- **filetype** searches for specific file types
  - ipod manual filetype:pdf
  - intext:"limited distribution" filetype:txt
  - "credit report request form" filetype:pdf site:ftc.gov
- Google's advanced search page allows only one file type per query & limits file type choices to only a handful of the thousands of known file types

- Bad news - Good news
  - Searching for info based on content creation date is difficult
    - No standards, no date info, or daily time stamping
    - Date created compared to date indexed
  - Consider 2 pages: one created 19990101, other 20090101
    - Search engine crawls and indexes both today
    - In terms of date indexed, the two pages are now the same
  - Search on how to search dates when you have time
  - "1 March 2009" OR "March 1, 2009" ...
    - Predict the future, include 2009..20XX in searches!
    - Don't forget about cultural dates (e.g. DPRK (Democratic People's Republic of Korea), Iran)



- info provides information about a site
- Google's "nuclear option"
  - `info:www.searchlores.org`

[Fravia's web-searching lore: Main entrance: Finding Information ...](#)

This totally advertisement-free **\*\*AND FROZEN\*\*** site teaches basic and advanced Internet search techniques for people interested in finding any kind of ...

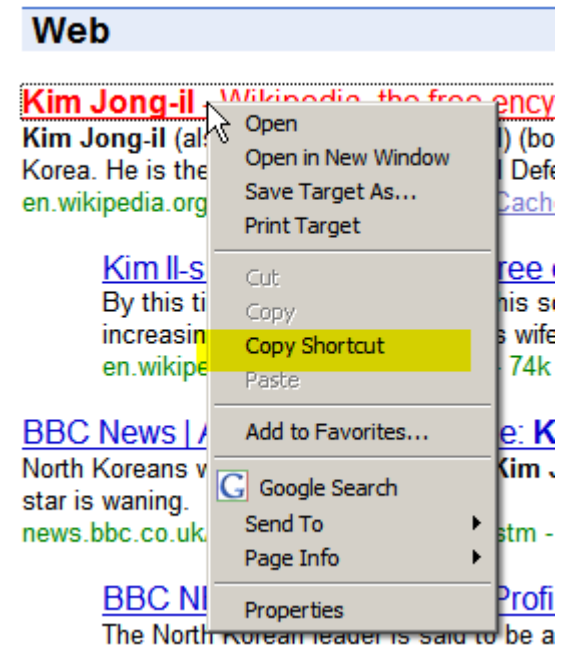
[www.searchlores.org/](http://www.searchlores.org/) - [Note this](#)

Google can show you the following information for this URL:

- Show [Google's cache](#) of `www.searchlores.org`
- Find web pages that are [similar to](#) `www.searchlores.org`
- Find web pages that [link to](#) `www.searchlores.org`
- Find web pages [from the site](#) `www.searchlores.org`
- Find web pages that [contain the term](#) "`www.searchlores.org`"

- Privacy issues
- No more mono-lingual manacles
- Supplemental Material (for home study)

- To protect privacy
  - Do not click on titles
  - Instead copy & paste URL in address line



- Use search engine translation tools
  - Translate "native" terms into "foreign" terms
    - Copy, paste, and search

Enter <b>text</b> or a webpage URL.	Translation: English » Thai
<input type="text" value="Pyongyang"/>	เปียงยาง

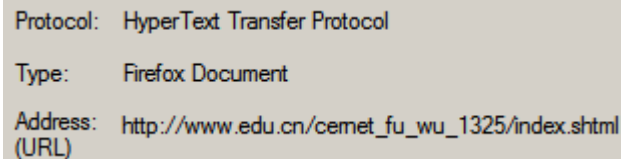
- Search, find, copy "foreign" terms
  - Translate "foreign" terms into "native" terms

Enter text or a webpage URL.	Translation: Chinese » English
<input type="text" value="教育在线"/>	Education Online

- Use language variants of same site creatively
  - Compare menus, figure captions, addresses, etc



- Right click and view "Properties" of links on both pages



The image shows a right-click context menu with the following information:

- Protocol: HyperText Transfer Protocol
- Type: Firefox Document
- Address: [http://www.edu.cn/cemet\\_fu\\_wu\\_1325/index.shtml](http://www.edu.cn/cemet_fu_wu_1325/index.shtml) (URL)

- View and use webpage source
  - Be prepared for a lot of hexadecimal jibberish; call IT guys



- Multi-lingual searches
  - Combine full, unknown vernacular with partially known English (or vice versa); no operators needed
  - Good for technical terms, organizations, individuals
    - "차기 전차" tank
  - Number of results can help with translation decisions
    - "차기 전차" future tank
    - "차기 전차" next generation tank
      - Even useful for English terms and phrases

# Advanced Search Form



## Advanced Search

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

Find results	with <b>all</b> of the words	<input type="text"/>	100 results ▾	<input type="button" value="Google Search"/>
	with the <b>exact phrase</b>	<input type="text"/>		
	with <b>at least one</b> of the words	<input type="text"/>		
	<b>without</b> the words	<input type="text"/>		
Language	Return pages written in	<input type="text" value="any language"/> ▾		
Region	Search pages located in:	<input type="text" value="any region"/> ▾		
File Format	<input type="text" value="Only"/> ▾ return results of the file format	<input type="text" value="any format"/> ▾		
Date	Return web pages first seen in the	<input type="text" value="anytime"/> ▾		
Numeric Range	Return web pages containing numbers between <input type="text"/> and <input type="text"/>			
Occurrences	Return results where my terms occur	<input type="text" value="anywhere in the page"/> ▾		
Domain	<input type="text" value="Only"/> ▾ return results from the site or domain	<input type="text"/> e.g. google.com, .org <a href="#">More info</a>		
Usage Rights	Return results that are	<input type="text" value="not filtered by license"/> ▾ <a href="#">More info</a>		
SafeSearch	<input checked="" type="radio"/> No filtering <input type="radio"/> Filter using <a href="#">SafeSearch</a>			

## Page-Specific Search

Similar	Find pages similar to the page	<input type="text"/>	<input type="button" value="Search"/>
		e.g. www.google.com/help.html	
Links	Find pages that link to the page	<input type="text"/>	<input type="button" value="Search"/>

- Useful information
- Tips and techniques
- Free stuff

- A few examples, can you find more?
  - Base/notation conversions
    - 1977 in roman numerals
    - (0x0F + 0x0F) in arabic numerals
  - Stock information
    - `stocks:intc`
  - Phonebook
    - John Doe VA

## Google recognizes pattern of specialized numbers, including:

- UPS, Fedex, and USPS tracking numbers
- VIN numbers
- UPC codes
  - 073333531084
- Map area codes
  - 571
- Track flights (powered by flightstats.com)
  - United 134



# Not all Queries are Logical

- Logical and illogical queries
  - site:cn site:com.cn
  - site:cn -site:com.cn
  - -site:cn site:com.cn
  - -site:cn -site:com.cn
    - What happened and why?
    - Which queries are redundant?
    - Which are logical? Which are not?

- How would you move from one to the other?
- Natural language query  $\Rightarrow$  Google advanced query
  - I want information from the OIE about the avian flu outbreak in 2005
  - I want to find blogs in Japan that talk about Iran's nuclear program
- Google advanced query  $\Rightarrow$  Natural language query
  - filetype:pdf site:edu intitle:"google search"
  - invisible OR hidden web inurl:blog -site:searchenginewatch.com

[www.google.com/options/specialsearches.html](http://www.google.com/options/specialsearches.html)

- University
  - Narrow search to specific school website
- U.S. Government
  - Search for content located on either U.S. federal, state and local government websites
- Apple
  - Search for Mac specific information, eliminate false hits
- Patent ([www.google.com/advanced\\_patent\\_search](http://www.google.com/advanced_patent_search))
  - Search full text of U.S. Patent files
- `inurl:cse inurl:coop site:google.com your_topic`

# Good Help is Hard to Find...

- Google

- <http://www.google.com/intl/en/help/refinesearch.html>
- <http://www.google.com/intl/en/options/>

- Yahoo

- <http://help.yahoo.com/l/us/yahoo/search/basics/basics-08.html>

- Windows Live!

- [http://help.live.com/help.aspx?project=wl\\_searchv1&market=en-us](http://help.live.com/help.aspx?project=wl_searchv1&market=en-us)


## www.pandia.com

### Comprehensive list of search engines and directories

- Custom search
- Local search
- Social search
- Reference
- Specialized searching
- Discussion groups
- Ezines
- Online books
- Video, MP3 and images
- Radio and podcasts
- RSS and blogs

All-in-one list of search engines and directories

### Pandia Powersearch



Search the Web:

Select search tool: Pandia Metasearch

**GO!** **HELP** [CLICK HERE FOR ADVANCED SEARCH FORM](#)

#### PANDIA

[Pandia Search Central](#)  
[Search Engine News](#)  
[SE Blogs and Sites](#)  
[Free Newsletters](#)  
[RSS web feed](#)

**Search tools:**  
[Powersearch All-in-One](#)  
[Plus Web Directory](#)  
[Metasearch](#)  
[Newsfinder](#)  
[Shopping Search](#)  
[Radio Search](#)  
[People Search](#)  
[Kids & Teens](#)

**On Web Searching:**  
[Search Tutorial](#)  
[Search Trends](#)

**On Search Ranking:**  
[SE Marketing Tutorial](#)

#### TOOLS FOR SEARCHING THE INTERNET

### All-in-One List of Search Engines

Pandia Powersearch is our all-in-one list of search engines and directories. Search the Web using the search form above, or select one of the categories below to find the best Internet search tools.


[Make This Page Your Home Page!](#)

THIS PAGE	OTHER PAGES
<a href="#">Search engines</a> <a href="#">Directories</a> <a href="#">Custom search</a> <a href="#">Local search</a> <a href="#">Databases</a> <a href="#">Social search</a> <a href="#">Reference</a>	<a href="#">Specialized searching</a> <a href="#">Discussion groups</a> <a href="#">Ezines</a> <a href="#">Online books</a> <a href="#">Video, MP3 and images</a> <a href="#">Online Radio and</a>


[Business](#)  
[Software](#)  
[Business](#)  
[Yellow Pages](#)  
[Translation](#)

[People and email](#)  
[News](#)  
[Shopping](#)  
[Search Engine Optimization](#)  
[Domain names](#)  
[Pay-per-click](#)


#### SPONSORS



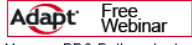
[Ready To Get New Customers?](#)



[The SEO dream team!](#)



[Outsource SEO/PPC](#)  
 \$11 - \$15 per hour



[Manage PPC Better - Just \\$199/mo.](#)

[YOUR AD HERE!](#)



Available as e-books via Safari database on [OpenSource.gov](http://OpenSource.gov)



***Google™pedia: The Ultimate Google Resource***  
Second Edition, 2007

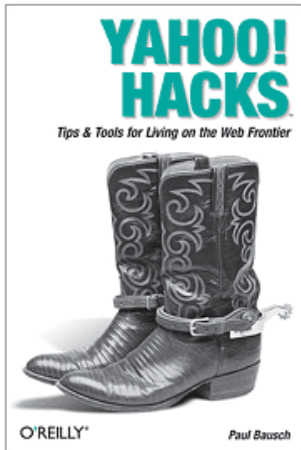


***Web Search Garage***  
by Tara Calishain, 2004



***Google® Power Tools Bible***  
by Ted Coombs; Roderico DeLeon, 2007

# Want to program Google hacks?



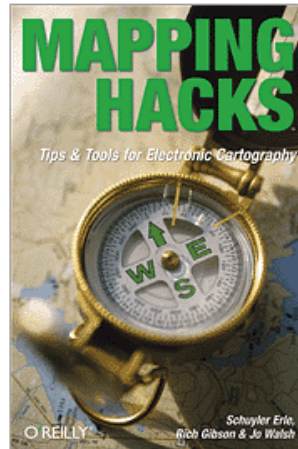
## Yahoo! Hacks

Tips & Tools for Living on the Web Frontier

By [Paul Bausch](#)

First Edition October 2005

Pages: 489 ([More details](#))



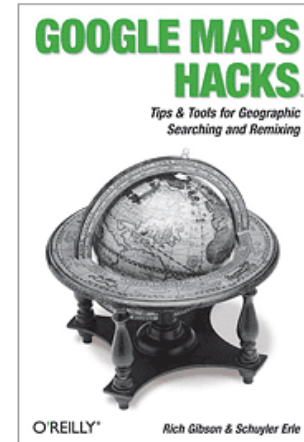
## Mapping Hacks

Tips & Tools for Electronic Cartography

By [Schuyler Erle](#), [Rich Gibson](#), [Jo Walsh](#)

First Edition June 2005

Pages: 564 ([More details](#))



## Google Maps Hacks

By [Rich Gibson](#), [Schuyler Erle](#)

First Edition January 2006

Pages: 366 ([More details](#))



## Google Hacks, Second Edition

Tips & Tools for Smarter Searching

By [Tara Calishain](#), [Rael Dornfest](#)

Second Edition December 2004

Series: [Hacks](#)

ISBN: 0-596-00857-0

479 pages, \$24.95 US, \$36.95 CA, £17.50 UK

All available in e-book  
format via Safari database  
on [OpenSource.gov](#)

# Search Engine Showdown

Compares and evaluates Internet search engines  
from the searcher's perspective.

- Search engine features chart
- Detailed search engine reviews
- An online newsletter
- Statistical analysis
- Search strategies



[www.searchengineshowdown.com](http://www.searchengineshowdown.com)

# Useful Blogs

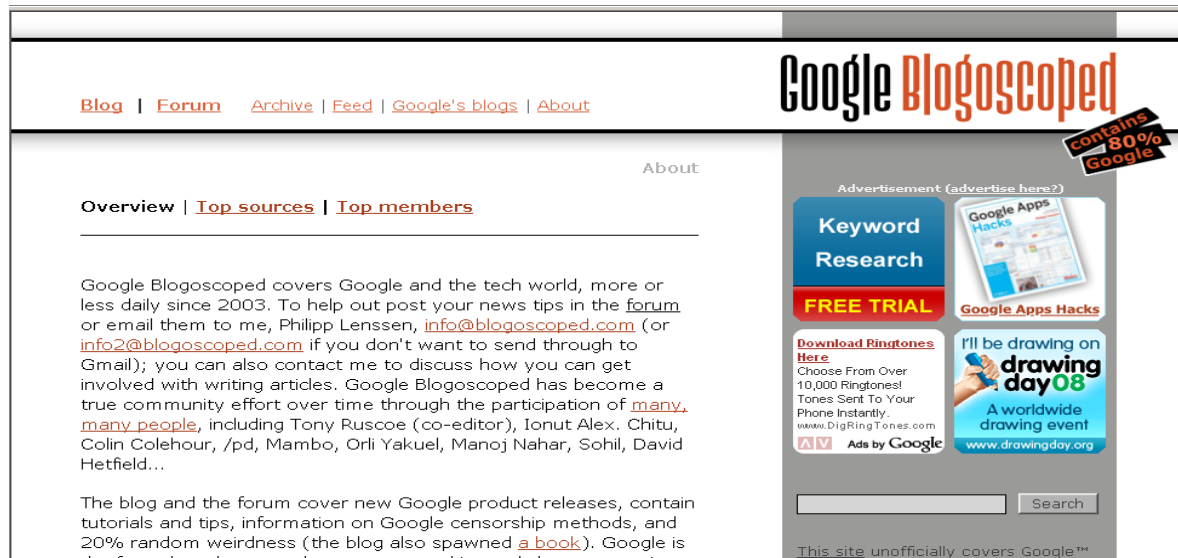
[www.researchbuzz.org](http://www.researchbuzz.org)

Check out the list of  
Google Custom Search  
Engines...



[www.blogoscoped.com](http://www.blogoscoped.com)

Great way to track  
Google product releases





- Parting thoughts
  - Start simple and refine your searches
  - FAVOR THE OBSCURE!!!
  - Be careful of spaces and typos
  - Too many operators or overuse of quotes can yield too few results
- Comments/Questions



# Exercises I

1. Find a list Live Search advanced search keywords. Note similarities and differences among Google's advanced operators
2. Use Live's **language** and **site** keywords to find Chinese Simplified pages registered to Thailand's domain. You will need to find Live Search language codes first.
3. Use Live's **language** and **site** keywords to find Chinese Traditional pages registered to Thailand's domain.
4. What language-CHS or CHT-has the larger presence in Thailand's domain?
5. Find a listing of Yahoo! search meta words. Note similarities and differences among Google's advanced operators.
6. Use Google to find information that Google has NOT yet indexed.

## Exercises II

1. Find other sites that discuss googling. Suggest using `site:edu`.
2. Find and explore some sites that compare search engines.
3. Go to [www.touchgraph.com/TGGoogleBrowser.html](http://www.touchgraph.com/TGGoogleBrowser.html) & experiment with, say, [www.cnn.com](http://www.cnn.com)
4. Go to [www.alexa.com](http://www.alexa.com); choose "Top Sites by Country"; choose a country; choose some website that interests you; explore that website using `site`; find a pattern in the URLs, then use `site & url` to explore that pattern.
5. Find [bbc.co.uk](http://bbc.co.uk) pages that don't start with `www`

## Exercises III

6. Do the queries below & explain the results
  - site:pref.\*.jp at Google
  - site:pref.\*.jp at Yahoo!
  
6. Translate Pyongyang into Hebrew and find images of North Korea using the Hebrew
  
8. Compare and explain; # results make sense?
  - inurl:media analysis \_\_\_\_\_
  - inurl:"media analysis" \_\_\_\_\_
  
9. Find www.state.gov in domains outside state.gov

10. `site:gov.cn` finds all Chinese government web sites: T or F and explain.
11. Compare and explain; # results make sense?
  - `site:cnn.com/news` \_\_\_\_\_
  - `site:cnn.com inurl:news` \_\_\_\_\_
12. Find comma separated value files from educational sites; view as HTML

13. `inurl:2009 site:usgs.gov`

- Enter above query at
  - `www.dogpile.com`
  - `www.mamma.com`
  - `www.vivisimo.com`
- What happened? Why?
- What do these results tell you?

14. Try `www.browsys.com`