

Googling to the Max - Exercises

Getting the most from: <http://www.google.com>

Google Toolbars (FREE) - highlight terms, remember searches, search within a site, block pop-ups, and more:

For Internet Explorer <http://toolbar.google.com>

For Mozilla (Netscape) <http://googlebar.mozdev.org>

#1 HOW GOOGLE "THINKS" IN DEFAULT MODE CRAFTING BASIC GOOGLE SEARCHES

AND automatically implied between terms

- all your terms somewhere
 - in text of pages
 - in pages that link to a result page
 - in other pages on the same site

Ex. 1: What is matched on?

1. Search Google for the keywords: **google page rank**
2. Click Cached: for the page titled "Google Technology" or "Google PageRank Calculator"
3. What is the explanation of the matching of your terms with these pages?

STEMS some words

- finds word with various endings
 - search **kite flying** and get matches on **kite, kites, kiting** and **flying, fly, flies**
- turn off with + or " " as in **+kite +flying** or **"kite flying"**

IGNORES common or "stop" words

- when this happens, a gray message appears below the search box in results telling you what was ignored
- turn off with + or " " just as for stemming

RANKING FAVORS pages with your words in PHRASES, CLOSE TOGETHER, and IN THE ORDER TYPED

Ex. 2: Word order and word choice matter

Compare the top results for searches in Google for these three sets of keywords:

grass snake	Mostly about a kind of snake. By stemming also matches snakes and grasses
snake grass	Many about a kind of grass, not present in 1 st search
snake in the grass	Most pages contain this expression, even though Google said in and the were ignored; also finds pages with grass snake and a few with snake grass
snake +in +the grass	Eliminates pages not containing in and the somewhere.
"snake in the grass"	Most specific and precise. Requires exact phrase in all pages.

OR searching requires capitalized OR

- Can be used between single words and phrases enclosed in quotes:
california OR oregon OR "pacific coast"
"global warming" OR "greenhouse effect"

No parentheses for nesting; keep OR searches simple

- 10 WORD MAXIMUM to search length
 - Workaround: Make phrases using the wildcard for whole words: * Replacing some words with *

Ex. 3: OR searches and hitting the 10-word maximum	
<ol style="list-style-type: none"> 1. What happens if you try to perform this search? ("global warming" OR "greenhouse effect") rise "sea level" (california OR "los angeles" OR "san diego" OR "san francisco") 2. Google ignores beyond 10 words: OR "san diego" OR "san francisco" 3. Substitute * for superfluous words that have little ambiguity: ("global warming" OR "greenhouse * ") rise "sea level" (california OR " * angeles" OR " * diego" OR " * francisco") = 10 words. OR not counted 	Parentheses ignored by Google.

#2 EXPLOITING GOOGLE "FUZZY" SEARCH OPTIONS

- PUNCTUATION THAT IS NOT IGNORED
 - Apostrophe (') : **peoples**, **people's**, and **peoples'** are searched as different words
 - Hyphen (-) : **same-sex** retrieves **same-sex**, **same sex** and **samesex**
 Always supply the - to search any word that might be used hyphenated
 - Accent marks in Roman-alphabet foreign languages : **é**léphant does not match **e**lephant (and vice versa)
 If searching in the language where the accent is common, the accents are not required for matches. Google assumes people writing web pages in that language may or may not put the accents in.
- SYNONYM SEARCHES (~)
 - Google will "think" of words with similar meaning: **~food** matches **recipes, nutrition, cooking**
~facts matches **information, statistics**
~help matches **guide, tutorial, FAQ, manual**
- SIMILAR PAGES (in results list) or command **related:[URL]**
 - Google will "think" of pages like the one you choose, by using links to and from the page, words in the page, and the importance of pages in links.
 - Uses: evaluate a questionable page by the links in and out
 find comparable pages when shopping or looking for a type of site
 broaden a search without thinking of words that might be in pages like the one you like

Ex. 4: Exploiting "FUZZY" Google options	
<ol style="list-style-type: none"> 1. Find pages about: <i>The one-child law of the People's Republic of China</i>, knowing that some people never use apostrophes properly: one-child law people's OR peoples "republic of china" 2. To find out, <i>How do praying mantises hear?</i> try searching: ~ears praying mantis 3. Find pages similar in focus to http://www.consumerwebwatch.org related:www.consumerwebwatch.org 4. Search the terms internet hoax In the results, find a page that seem intriguing, click <u>Similar pages</u>. How is the second page related to the first? 	

- NUMBER RANGE allowed in searches (.. without spaces)
 - Follow search terms with beginning and ending numbers, separated by two periods.
 - Can be used one sided, as *less than* or *greater than*

<p>Ex. 5: Number range searches</p> <ol style="list-style-type: none"> 1. Find pages mentioning Babe Ruth between 1921 and 1935 "babe ruth" 1921..1935 2. Find digital cameras priced below \$300 with resolution of 4.5 to 5 megapixels (MP) digital camera ..\$300 4.5..5 megapixels OR MP

#3 SETTING LIMITS IN GOOGLE SEARCHES

- Require search terms to occur in specific parts of web pages
 - **intitle:** Words must occur in the official <title> field in the head part of the page
 - **site:** Limit to a site or domain (first part of the URL, before the first /)
 - **inurl:** Require terms to occur anywhere in URL (URL punctuation ignored)
 - Can be combined with each other and with other search terms

<p>Ex. 6: Limiting to parts of web pages</p> <ol style="list-style-type: none"> 1. Find pages primarily about mileage in hybrid cars: intitle:mileage "hybrid cars" 2. Find pages about President Bush and either global warming or greenhouse effect: bush intitle:"global warming" OR intitle:"greenhouse effect" 3. Is there anything on this subject from the official Whitehouse site? site:www.whitehouse.gov "global warming" OR "greenhouse effect" 4. Find pages containing sheet music written to commemorate the assassination of Abraham Lincoln site:memory.loc.gov lincoln "sheet music" Searches within all Library of Congress's American Memory Project pages that are in Google. inurl:lincoln "sheet music" Finds pages with "sheet music" in them and with "lincoln" somewhere in the URL.
--

- Limit to pages updated within a range of dates: **daterange:[Julian date-Julian date]**
 - Matches on the date Google last visited the page without detecting any changes; Google stores the date of each visit (visible in Cache), but does not change the date searched in daterange: if the page has not been altered.

<p>Ex. 7: Limiting to a range of dates visited with content changed:</p> <p>Find pages about the World Trade Center that were posted to the web on September 11 or 12, 2001, and have not been changed since: "world trade center" daterange:2452163-2452164</p> <p>To obtain Julian dates (number of days since noon Jan. 1, 4713 B.C.): use http://www.faganfinder.com/google.html use Calendar Converter at http://www.fourmilab.to/documents/calendar</p>

- Limit to type of document or file extension: **filetype:**
 - Google has many types of files besides HTML pages (doc, xls, wpd, ppt and more)
 - View as HTML link in results with these file types lets you see the document without the application and avoids any virus or worm they could contain

Ex. 8: Limiting to type of document or file:

1. Find fact sheets and reports (not just web pages) about the death-rate for malignant mesothelioma in the U.S.
filetype:xls OR filetype:pdf death-rate malignant mesothelioma "united states"
2. Find PowerPoint presentations from reputable universities in the U.S. about the impact in China of its one-child policy
filetype:ppt site:edu china one-child policy

- **Google Advanced Search** does not allow *intitle:*, *inurl:*, or more than one *site:*, and has few date range options.
- **Pages from within a foreign country by building on the URL of a Google search result**
 - Uses information about where the webpage server is located, as well as country coding; more comprehensive than what is available a http://www.google.com/language_tools
 - Search on the terms you want. In your browser's Address or URL box, append to the result (without spaces):
&restrict=countryXX with XX being the CAPITALIZED two-letter country code
 To find country codes: Search Google for **TLD [country name]** or
 List of all country codes available at http://en.wikipedia.org/wiki/List_of_Internet_TLDs

Ex. 9: Pages from within a country:

Find recipes for authentic Moroccan cuisine, from within Morocco

Step 1: perform the search:
recipes OR cuisine morocco OR moroccan

Step 2: append to the resulting URL in the Address box **&restrict=countryMA**
<http://www.google.com/search?q=recipes+OR+cuisine+morocco+OR+moroccan&restrict=countryMA>

#4 HANDY GOOGLE TOOLS AND SHORTCUTS

- **DEFINITIONS, THESAURUS, and ENCYCLOPEDIA LOOK-UPS**
 - **define:[your words]** Finds web pages about your word, Google-ranked, and often reliable
 - **define [your words]** Finds web definition for the first result with link to other web definitions, also finds web pages with your words
 - **search and click your word in statistics bar** Access to published dictionary sources, with synonyms, antonyms, and other information. Click THESAURUS button if needed.

Ex. 10: Definitions and more:

1. Compare the results from the searches:
define internet
define:internet
2. Locate dictionary definitions and thesaurus terms (synonyms, antonyms, etc.) using the link in the blue statistics bar after you search:
scholarship

☐ TRANSLATIONS

- Click Translate the page after a page in a foreign language
- Paste a URL or a piece of text at http://www.google.com/language_tools

Ex. 11: Translating:
1. Go to Google Advanced Search - www.google.com/advanced_search
2. Limit to PAGES IN FRENCH and search jacques chirac avec george bush
3. Click <u>Translate this page</u> (appears with most pages in languages not in your search interface)
4. You can also translate by pasting a URL or text - www.google.com/language_tools

☐ CALCULATOR and CONVERSIONS - parentheses accepted to sequence and clarify expressions

- **Mathematical functions** (+, -, *, /, % of, *n*th root of, sqrt(*nn*), ^ for exponentiation)
- **Advanced math** (Trig: sin, cos, tan, sec, csc, cot, etc.; Inverse trig: arcsin, arcos, arctan, etc.; Hyperbolic trig: sinh, cosh, tanh, etc.; Logarithms; Exponential functions; Factorials; and more)
- **Many mathematical constants** (pi, imaginary numbers, and more)
- **Units of measure and conversions**
 - . MASS: kilograms or kg, grams or g, grains, pounds or lbs, carats, stones, tons, tonnes, etc.
 - . LENGTH: meters or m, miles, feet, inches, Angstroms, cubits, furlongs, etc.
 - . VOLUME: gallons, liters or l, bushels, teaspoons, pints, drops, etc.
 - . AREA: square miles, square kilometers, acres, hectares, etc.
 - . TIME: day, seconds or s, years, centuries, sidereal years, fortnights, etc.
 - . ELECTRICITY: volts, amps, ohms, henrys, etc.
 - . ENERGY: Calories, British thermal units or BTU, joules, ergs, foot-pounds, etc.
 - . POWER: watt, kilowatts, horsepower or hp, etc.
 - . INFORMATION: bits, bytes, kbytes, etc.
 - . QUANTITY: dozen, baker's dozen, percent, gross, great gross, score, etc.
 - . NUMBERING SYSTEMS: decimal, hexadecimal or hex, binary, roman numerals, etc.
- **Physical constants** (atomic mass units or amu, Avogadro's number, Boltzmann constant, Faraday constant, gravitational constant, mass of a proton, mass of each planet and of the sun, permeability of free space, etc.)

Ex. 12: Calculating and converting:
1. How many carats in a pound? 1 lb in carats
2. How many square feet in 2.5 hectares? 2.5 hectares in square feet
3. Value of 1555 in hexadecimal? 1555 in hex
4. The mass of an electron? mass of an electron
5. How many bits in a 2.5 gigabytes? 2.5 gigabytes in bits

☐ SHORTCUTS to many other kinds of publicly available information

- Area codes (U.S.), defined by maps - type **3 digit code**
- Airport conditions - type **airport code** or **city** and the word **airport**
- Flight tracking - type the word **airline** and the **flight number**
- Maps - type an **address**
- Stock info - type **one or more NYSE, AMEX, NASDAQ, or mutual fund codes**
- White pages - type **rphonebook: name address**
- Yellow pages - type **bphonebook: name address**
- Zip codes - type a **zip code**

☐ STANDARD NUMBER LOOKUPS

- Parcel tracking for UPS, FedEx, USPS - type the **number**
- US patent lookup - type the word **patent** and the **number**
- UPC product identification - type the **code**
- VIN vehicle info, history - type the **number**

Ex. 13: Shortcuts and standard number lookups:

1. What part of the U.S. is in the zip code 87455
87455 Look below the map at the town and state.
2. Are there currently any delays at O'Hare airport in Chicago?
chicago airport or **ord airport**

UNIVERSITY SEARCHES

- Conduct a search within the website of a university - you specify the topic
- <http://www.google.com/options/universities.html> has long list of University links to start searching from

Ex. 14: University searches:

Go to Google Advanced Search. At the bottom, click on Universities.
Choose any university from the list and search for information on:
graduate admissions

#5 THE BEST OF GOOGLE'S FAMILY OF DATABASES

Images

- Largest images database - 800+ million

News

- 4,500 news feeds, 30 days of news

Shopping Databases

- Froogle - shopping sites from Google web page database + merchant-supplied machine-readable catalogs
- Catalogs (Google Labs)- scanned catalogs, text-searchable + web pages

Local

- Google web pages + several yellow page directories
- For "finding businesses and services"
. Two search boxes: enter a place or address and a topic

Groups

- Google **Groups** - Usenet Newsgroups back to 1981
. Unevaluated opinions on many topics
. Can limit to a specific forum
- Google **Groups2** (Google Labs) - old Groups + create new groups

More in Google Labs


- Google **Sets** - generate items in a set you start
- Google **Webquotes** - find out what other websites say about a site
- More at labs.google.com

Ex. 15: Sampling Google databases:

- At the Google homepage (<http://www.google.com>), click on **more>>**
1. Compare shopping for **diamond tiara** in Froogle and Catalogs.
 2. Use Local Search to search for **temporary employment OR jobs** in your home town.
 3. Go to Advanced Image Search. Find **full color** pictures in **JPG** format of **mona lisa**.

#6 GOOGLE PRINT and GOOGLE SCHOLAR

Two of Google's initiatives rather directly involve libraries, published information, and access to published information, both on web pages and in print. There seems to be quite a bit of confusion between these two services, and so I am organizing this discussion as a two-column comparison.

	Google Print	Google Scholar
Google's goal	To "organize the world's information and make it universally accessible and useful" focused on content of books	To "search specifically for scholarly literature" in many formats
How to use	Through the regular Google search boxes (google.com) Books icon  indicates links to content in print.google.com	Through special a search box at scholar.google.com
What's included	<ul style="list-style-type: none"> <input type="checkbox"/> Publisher-supplied publications; also authors of self-published and out-of-print publications can submit their works as the publisher <input type="checkbox"/> Publications supplied by selected libraries (Harvard, Stanford, U of Michigan, Oxford, NYPL), both copyright and public domain <input type="checkbox"/> Google scans, digitizes, and indexes words in each book for free (no charge, no profit) <input type="checkbox"/> Mostly books, but some articles (primarily <i>Variety</i>, <i>Library Journal</i>, <i>School LJ</i>, <i>Publishers Weekly</i>, and technical, scientific, and pharmaceutical publications) 	<ul style="list-style-type: none"> <input type="checkbox"/> Mostly articles, but some citations to books <input type="checkbox"/> Full-text of online articles found by Google's robot-like crawlers from academic publishers, professional societies, preprint repositories, and scholarly articles posted on the web <input type="checkbox"/> Citations to articles and books found in lists of references in publications found by Google's robot-like crawlers
What you get	<ul style="list-style-type: none"> <input type="checkbox"/> Scanned full text to search and browse - no printing of pages or images <input type="checkbox"/> Snippets or excerpts of copyright books when publisher has not permitted full text online <input type="checkbox"/> Copy of table of contents and index <input type="checkbox"/> Link to publisher, bookstores, and libraries if a scanned library copy <input type="checkbox"/> "About this book" resembles brief catalog or BIP entry <input type="checkbox"/> Paid advertisements related to your search, as in Google web search 	<ul style="list-style-type: none"> <input type="checkbox"/> Full text if freely available or if provided by the publisher with permission <input type="checkbox"/> Citation or abstract (not full text) if protected by copyright <input type="checkbox"/> Links to citations to most articles (like a "who cites whom") <input type="checkbox"/> For books, "Library Search" tries to locate a library near you that owns the book. This is only a partial subset of library holdings, not a complete reflection of holding libraries
Search tips	<ul style="list-style-type: none"> <input type="checkbox"/> Include the word "books" in your search <input type="checkbox"/> Include site:print.google.com in your search <input type="checkbox"/> <i>For articles only:</i> inurl:articleid site:print.google.com your search terms including journal title if known <input type="checkbox"/> <i>For books only:</i> inurl:isbn site:print.google.com your search terms including journal title if known 	<ul style="list-style-type: none"> <input type="checkbox"/> Go to scholar.google.com or Scholar Advanced Search http://scholar.google.com/advanced_scholar_search <input type="checkbox"/> Enter author name, publication name, and/or subject words <input type="checkbox"/> If full text is not available online, the best bet is to search a library's catalog or article database services directly <input type="checkbox"/> Sometimes the publication can be purchased online
More information	About Google Print http://print.google.com	About Google Scholar http://scholar.google.com/scholar/about.html