

GOOGLE HACKS™

**2nd
Edition**
Covers Gmail

Tips & Tools for Smarter Searching



Tara Calishain & Rael Dornfest

*With a new foreword by
Craig Silverstein, Director of Technology, Google*

O'REILLY®

HACK
#21

Like a Version

Gather a list of what Google thinks are synonyms for a keyword you provide.

The Google `~` synonym operator [**“Special Syntax” in Chapter 1**] widens your search criteria to include not only the specific keywords in your search, but also words Google has found to be synonyms of, or at least in some way related to, your query words. So while, for example, `food facts` may only match a handful of pages of interest to you, `~food ~facts` seeks out nutrition information, cooking trivia, and more. And finding these synonyms is an entertaining and potentially useful exercise in and of itself. Here’s one way...

Let’s say we’re looking for all the synonyms for the word “car.” First, we search Google for `~car` to find all the pages that contain a synonym for “car.” In its search results, Google highlights synonyms in **bold**, just as it highlights regular keyword matches. Scanning the results (the second page is shown in [Figure 2-1](#)) for `~car` finds `car`, `cars`, `motor`, `auto`, `BMW`, and other synonyms in boldface.

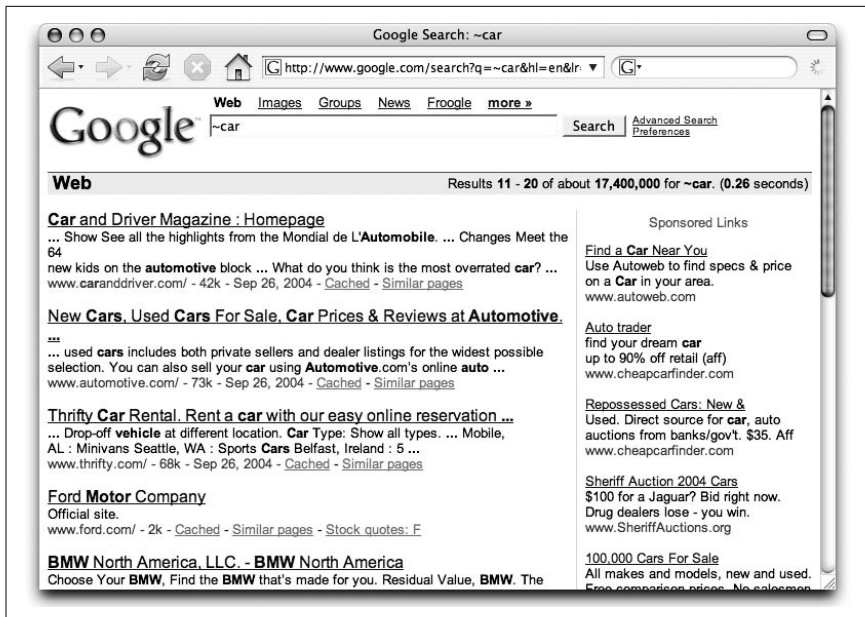


Figure 2-1. `~car` turns up bolded synonyms in Google search results

Now let’s focus on the synonyms rather than our original keyword, “car.” We’ll do so by excluding the word “car” from our query, like so: `~car -car`. This saves us from having to wade through page after page of matches for the word “car.”

Once again, we scan the search results for new synonyms. (I ran across automotive, racing, vehicle, and motor.)

Make a note of any new bolded synonyms and subtract them from the query (e.g., ~car -car -automotive -racing -vehicle -motor) until you hit Google's 10-word limit [["The 10-Word Limit" in Chapter 1](#)], after which Google starts ignoring any additional words that you tack on.

In the end, you'll have compiled a goodly list of synonyms, some of which you'd not have found in your typical thesaurus thanks to Google's algorithmic approach to synonyms.

The Code

If you think this all sounds a little tedious and more in the job description of a computer program, you'd be right. Here's a short Python script to do all the iteration for you. It takes in a starting word and spits out a list of synonyms that it accrues along the way.



You'll need the [PyGoogle \[Hack #98\]](#) library to provide an interface to the Google API.

```
#!/usr/bin/python
# Available at http://www.aaronsw.com/2002/synonyms.py
import re
import google # get at http://pygoogle.sourceforge.net/
sb = re.compile('<b>(.*?)</b>', re.DOTALL)
def stripBolds(text, syns):
    for t in sb.findall(text):
        t = t.lower().encode('utf-8')
        if t != '...' and t not in syns: syns.append(t)
    return syns
def findSynonyms(q):
    if ' ' in q: raise ValueError, "query must be one word"
    query = "~" + q
    syns = []

    while (len(query.split(' ')) <= 10):
        for result in google.doGoogleSearch(query).results:
            syns = stripBolds(result.snippet, syns)

    added = False
    for syn in syns:
        if syn in query: continue
        query += " -" + syn
        added = True
        break
```

```
        if not added: break # nothing left

    return syns
if __name__ == "__main__":
    import sys
    if len(sys.argv) != 2:
        print "Usage: python " + sys.argv[0] + " query"
    else:
        print findSynonyms(sys.argv[1])
```

Save the code as *synonyms.py*.

Running the Hack

Call the script on the command line [[“How to Run the Hacks” in the Preface](#)], passing it a starting word to get it going, like so:

```
% python synonyms.py car
```

The Results

You'll get back a list of synonyms like these:

```
['auto', 'cars', 'car', 'vehicle', 'automotive', 'bmw', 'motor', 'racing',  
'van', 'toyota']
```

—Aaron Swartz