Super Searcher Strategies

Mary Ellen Bates  March 28, 2017
Reluctant—Entrepreneur.com
Tweeting this?

#CILDC

@mebs
Google search tips & tricks
Google’s RankBrain kicks in

Machine learning
Boolean isn’t Boolean any more
“We think you really meant this…”
Less emphasis on text-search features
More emphasis on your search history
Impact on search pros

“Everything you know is wrong”

Forget finality, certainty, repeatability

Creativity counts
Don’t be too clever

Too many terms = lower results
Don’t use long OR statements
Use Google’s auto-synonyming
Graze, don’t just dig deep
Try multiple approaches

For terms you want closely related, try all these:

word1 word2
word1–word2
“word1 word2”
Inexact "exact" searching

Verbatim
Search tools -> All Results -> Verbatim

Intext:
To avoid hits from anchor text links

“word”
Use * strategically

* replaces any term in URL

For innovation initiatives at NASA:

```
site:*.nasa.* inurl:innovation
```

For finding email addresses with middle initials

“mary.*.bates@batesinfo.com”
Google Web ≠ Google Scholar

Search both Web and Scholar to catch trends, experts

Use Scholar’s Advanced Search feature
Custom Search Engine

Google still does custom search!
Favor/require/suppress specific URLs
Append search terms
Help your clients search better!
Fight fake information!

cse.google.com
### Reluctant-Entrepreneur.com

#### Google's "My Activity"

### myactivity.google.com

#### Creepy or useful?

<table>
<thead>
<tr>
<th>ITEMS</th>
<th>MAPS</th>
<th>SEARCH</th>
<th>HELP</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:25 PM</td>
<td>Maps and Search</td>
</tr>
<tr>
<td></td>
<td>9 items</td>
</tr>
<tr>
<td>3:24 PM</td>
<td>google.com</td>
</tr>
<tr>
<td></td>
<td>Searched for <em>What is Uber's market strategy</em></td>
</tr>
<tr>
<td></td>
<td>Details · Search</td>
</tr>
<tr>
<td></td>
<td>20 items</td>
</tr>
<tr>
<td>3:18 PM</td>
<td>mapsengine.google.com (4 times)</td>
</tr>
</tbody>
</table>
OTHER search engines
Bing

Always my *second* choice

Different index, different ranking

**linkfromdomain:**

See pages linked *from* a URL

**linkfromdomain:**lifehacker.com todo

finds good “to-list” tools
Bing

Find pages that link to a file type with contains:
contains:mp3
contains:xls
contains:pdf
Peekier.com

Anonymous search engine

Mostly Bing, and its own crawl

Preview images of the results
"fuel insecurity" biofuels
Millionshort.com

‘What haven’t you found?’™
Filter by site popularity
Eliminate ‘click-bait’

Essential for people-searching, other long-tail niche searches
Diabetes Type 2 Recipes - Cooking For Diabetics ... 
http://cookingfordiabetics.org/diabetes-type-2-recipes/
Type 2 diabetes (sometimes called mature onset diabetes) is the most common form of diabetes, accounting for 90% of cases diabetes. This disease affects nearly 17 ...

Recipes - Diabetes Queensland
Diabetes Queensland provides you with healthy recipes to assist you in maintaining a healthy lifestyle with regular exercise. Submit your own recipe.

Living with Type 2 Diabetes - Easy Recipes for Diabetics ...
Type 2 diabetes, find the easy to prepare, effective day to day diabetic recipes, cheaper foods diet plans

Triple chocolate brownies - Diabetes Queensland
Diabetes Queensland provides you with healthy recipes to assist you in maintaining a healthy lifestyle with regular exercise. Submit your own recipe.

Recipes For Diabetes Type 2
Search the web, save the planet

Ecosia.org
Plants trees in Africa with its ad revenue
(6.5 million 🌳 so far!)
Powered by Bing
Other search tricks
A case for case studies

Need examples of who’s doing something well?

Look for CASE STUDIES on the topic

See Marydee Ojala’s Dollar Sign column in Jul/Aug 2016 Online Searcher
EVERY CRS report!

everycrsreport.com

Nonpartisan perspective

Full-text searchable

8,000 reports and counting
Clients don’t know what to ask for

What other data sets would be helpful?

What is your client not asking you for?
Knowing when 0 means 0

WHY did you get no results?
- Typos, transposed letters
- Wrong logic or syntax
- Not using common nomenclature
- Unquestioned assumptions
- Wrong database/file/index
Knowing when 0 means 0

Tell a story

Do the gaps correlate to an event?
What related info did you find?
What do you recommend next?
Use link searches for insight

Unfamiliar site?
Validate by seeing who links to it
Do linking sites show a bias or PoV?
Review outgoing links; where to?
Link searching

Gigablast syntax is \texttt{link:the.url}

Or try openlinkprofiler.org
Remember the deep web

Use search engines for leads:

*Keywords* (database OR dataset OR archive)

*Keywords* (portal OR resources OR “online tool”)
Finding deep web content

Start with one known source
Association, NGO, non-profit, library or museum, etc.

Look for govt portals
data.gov, usa.gov/statistics, etc.

Then use Bing’s linkfromdomain:
Reverse image search

Is anyone using our photos?

ID useful chart, person

Then look for other copies of that image
Reverse image search

Search by image URL or drag/drop an image
Google Image Search, click the 📷 Tineye.com
More image searching

Use image search to find reports!

Example:

search Google Images for "dog food"
"market share"

Skim/mouse-over results
Searching with images

Notice that petfoodindustry.com is a good source

Then search the web for

“market share” 2016..2018
site:petfoodindustry.com
Get leads for experts

site:twitter.com/*/lists inurl:librarians

Gets Twitter lists about librarians
ID similar orgs in LinkedIn

- Nordic Heritage Museum
  - Museums and Institutions
  - 11-50 employees

- OVPM / OWHC / OCPM
  - Government Relations
  - 11-50 employees

- HUGH STUART EXPLORATION CONSULTANTS LTD
  - Mining & Metals
  - 2-10 employees

- American Institute for Conservation of Historic and Artistic Works
  - Nonprofit Organization Management
  - 11-50 employees

- UNESCO
  - International Affairs
  - 1,001-5,000 employees
Curate open access resources

Open-access journals: doaj.org
Open-access books: doabooks.org
Open-access repositories: opendoar.org
Curate open access resources

Public Library of Science: plos.org

OpenGrey (Grey Lit in Europe).opengrey.eu

BASE – Bielefeld Academic Search Engine.base-search.net
Can’t ID a phone number?

Look it up in Facebook!
# Privacy Settings and Tools

<table>
<thead>
<tr>
<th>Who can see my stuff?</th>
<th>Who can see your future posts?</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Review all your posts and things you're tagged in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limit the audience for posts you've shared with friends of friends or Public?</td>
<td></td>
</tr>
<tr>
<td>Who can contact me?</td>
<td>Who can send you friend requests?</td>
<td>Everyone</td>
</tr>
<tr>
<td>Who can look me up?</td>
<td>Who can look you up using the email address you provided?</td>
<td>Everyone</td>
</tr>
<tr>
<td></td>
<td><strong>Who can look you up using the phone number you provided?</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>This applies to people who can’t already see your phone number.</td>
<td></td>
</tr>
</tbody>
</table>

| Everyone |
Preview shortened URLs

checkshorturl.com

Type in short URL

See the expanded version, screenshot, description
<table>
<thead>
<tr>
<th><strong>Long URL</strong></th>
<th><a href="https://reluctant-entrepreneur.com/2017/02/04/how-to-make-yourself-competition-proof/">https://reluctant-entrepreneur.com/2017/02/04/how-to-make-yourself-competition-proof/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delay</strong></td>
<td>2.14 second(s)</td>
</tr>
<tr>
<td><strong>Short URL</strong></td>
<td><a href="http://owl.li/1TNE308GEC5">http://owl.li/1TNE308GEC5</a></td>
</tr>
<tr>
<td><strong>Redirection</strong></td>
<td>301</td>
</tr>
<tr>
<td><strong>Search long URL on</strong></td>
<td>Yahoo</td>
</tr>
<tr>
<td><strong>Check if safe on</strong></td>
<td>Web Of Trust</td>
</tr>
<tr>
<td><strong>Title</strong></td>
<td>How to make yourself competition-proof</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>I was chatting with a friend the other day who, among other things, is a professional wallpaper installer. A local hotel had contacted her and asked for a bid to redo all the wall coverings in their halls. She knew that her</td>
</tr>
</tbody>
</table>
PrintFriendly.com

Creates a printer-friendly version of a web page

Edit out images, extraneous matter, change font size

Browser plug-in, too
Google kills SHA-1 with successful collision attack

By Fahmida Y. Rashid

It's official: The SHA-1 cryptographic algorithm has been "SHAtered." Google successfully broke SHA-1. Now what?

After years of warning that advances in modern computing meant a successful collision attack against SHA-1 was imminent, a team of researchers from Google and Centrum Wiskunde & Informatica (CWI) in the Netherlands have successfully developed the first successful SHA-1 collision. In practical terms, SHA-1 should not be relied upon for practical security.

Modern cryptographic hash functions depend on the fact that the algorithm generates a different cryptographic hash for every file. A hash collision refers to having two separate files with the same hash. The fact that cryptographic weaknesses in SHA-1 make certificates using the SHA-1 algorithm potentially vulnerable to collision attacks is well-known. The National Institute of Standards and Technology deprecated SHA-1 more than five years ago, and experts have been long urging organizations to switch to stronger hash algorithms. Up until now, the only thing going for SHA-1 was the fact that collision attacks were still expensive and theoretical.

No longer, as the Google-led research team has developed a method that let them generate two PDF files with different content but generating the same SHA-1 hash. While the collision attack is still expensive, the "SHA-1 shattered" attack is no longer theoretical, which means the attack is within the reach of anyone motivated enough and with deep enough pockets.

“We started by creating a PDF prefix specifically crafted to allow us to generate two documents with arbitrary distinct visual contents, but that would hash to the same SHA-1 digest," the team from Google and CWI wrote in a blog post. "We were able to find this collision by combining many special cryptoanalytic techniques in complex ways and improving upon previous work."

However, it's worth noting that forging digital certificates will remain difficult thanks to the new CA/Browser Forum rules that require 20 bits of randomness to be added to digital certificate serial numbers.

SHA-1 is dead; act accordingly

Even though many organizations have been working on migrating to SHA-2 for the past year, the switchover isn't 100 percent complete, which means organizations who have not yet finished (or started!) their switchover are now at risk. Attackers now have proof collision attacks are possible, and under Google's disclosure policy, the code that would allow attackers to create these PDF documents will be publicly available.

[18 surprising tips for security pros. Security Report newsletter.]
Find Mary Ellen Bates at:

mbates@batesinfo.com
Twitter: @mebs
LinkedIn: maryellenbates
Facebook: maryellenbates
Reluctant-entrepreneur.com