

Live Web Search: Hacking the Blogosphere

Mary Hodder
CEO: Dabble, Inc.

mary@dabble.com

Blog: [napsterization \(napsterization.org/stories\)](http://napsterization.org/stories)

Live Web Buzz: Measuring the Blogosphere

Doc Searls:
'the live web'



Search: static web

- Google was created for the static web
 - Pagerank is a function of people's interest on one site, in another site, as expressed in links.
 - Google's genius was equating those links to social gestures that had value.
 - Key word search returns the 'most relevant' results based upon whomever has the most pagerank.

Search for 'mary hodder' in Google

Yields:

- 'napsterization'
- 'biplog'
- 'speaking engagements'
- 'dabble'
- 'hodder.org'
- Matches for 'mary hodder'

Search for 'napsterization' in Google

Yields:

- 'mary hodder'
- The napsterization blog itself
- Uses of the word, napsterization

Search for 'mary hodder' or napsterization in blog search

- Yields key word matches
 - Reverse chron order
- Link search: yields all those who have linked to url
 - Reverse chron order
 - Can resort by 'Authority' in some services

Static webpage search

- Static Webpages: Example webpage from a bank.
 - Incoming links 'define the bank' referring to it as an entity, but the bank doesn't change it's pages.
 - Outgoing links from bank site rarely change.
 - Don't really want sites like that to update a lot; useful in static nature

Metrics, Communities and Weighting Blogs

Why do we care?

What is it?

What does it mean?

How do we do it?

Live webpages: what and how

- Live web pages: rapidly updated pages such as Flickr photo sets, or blogs, often with lots of links and references in new posts.
- The constant activity produces more links and references in and out of blogs, as conversations between creators appear in threads
 - Old blog entries scroll down and off blogs, but in Google sees them all as subpages
 - Page rank makes no distinction because it defines pages as 'any page' instead of 'top pages' as technorati, or pubsub do when they 'age' things out of the system..

Metrics of Blog Search

Examples of those who rely on link counts:

- Technorati
- Bloglines

“authority” vs.
“influence”

The screenshot shows a web browser window displaying a search results page on Technorati. The search query is "jerry reynolds" and the results are sorted by "All Languages". The page shows 62 posts about "jerry reynolds" in the last 89 days. The top result is titled "Chevy poised to pass Ford" and is from Universal Automotive Database. The second result is titled "Spammer Tries To Push Delete On Spaming Past Jerry Reynolds" and is from MemeBattle. The third result is titled "This makes me laugh..." and is from Boing Boing. The page also features a sidebar with "Sponsored Links" for a Springfield, MO Attorney, "Most Popular" news items, and "Photos" from Buzznet & Flickr.

Live Web Examples

Bloglines: 928,815,007 Articles Indexed* as of 8/05

Ice Rocket: no info

Sphere: no info

Technorati: 50 m + blogs indexed

Del.icio.us: tags, urls and people feeds

Flickr: Photos, tags and sets/groups by 2.5m
people with 100 m + photos

Alexa: visits to a site based on toolbar
installs/visits

Link metrics used to weigh bloggers

- Technorati: counts all inbound link sources for last 6 months
- Bloglines: count of all links for all time *and* subscription counts w/in service
- Sphere: uses links plus other weights for relevance
- Feedburner: counts all served subscribers *only for blog owners

Subscription metrics

Example: Boing Boing * as of August, 2005

- In Bloglines, 26,856 subscribers
- In Feedburner, 1.2 million subscribers to feed
- Difference: Feedburner shows all subscriptions where as Bloglines shows just those who use Bloglines reader and subscribe



Possible other metrics

Find these yourself for a better picture of a blogger:

- **Inbound links: to post url**
- **Inbound links: to blog url**
- **Comments to posts**
- **Frequency of blogging a topic**
- **Subscribers to RSS feeds and blogs**
- **Incoming Traffic Links**
- **Outbound blogroll**
- **Tagged urls**

Live Web Search results

- Typically returned in reverse chronological order or factored in terms of time
- Sphere: can order by date or by relevance (metric based on links and key word analysis)
- Technorati: can reorder by 'authority' (defined as inbound link count)

Results of mixing live web with Google:

- Google: has lack of awareness of time about publishing of site
 - Good: information continues to be available
 - Bad: liveness is lost in results but activity can overshadow regular static web search
- Awareness of links and activity
- Result: Live web activity can overwhelm static web

Example: 'Jerry Reynolds'

- Spammer from 1990's
 - Largest Usenet spammer in 1997
 - Owner of two sites: netzilla and sexzilla
 - His spam helped, with other spam, to kill Usenet over several years
 - Spam fighters tried to stop it but failed
- Wants this information removed
 - Using C&Ds to remove this documentation from Google
- Why? Google owns Usenet postings.. They come up in searches

Before blogging about Jerry Reynolds

- 13 pages back in Google Results
- Unknown for years by anyone in technology journalism
 - Exception: one article in local paper in North Dakota

Blogging of Jerry Reynolds story

- BoingBoing
- Splatterings
- Memeblog

The screenshot shows a web browser window with the address bar displaying 'http://doc.weblogs.com/2005/11/07#spammerSlappsArour'. The page title is 'The Doc Searls Weblog : Monday, November 7, 2005'. The main content area features a post titled 'Spammer SLAPPs around critics' dated 'Monday, November 7, 2005'. The post text includes: 'Mary Hodder: Jerry Reynolds, (former) Spam Slapp Suits Trying to Squash Free Speech by Ways. SLAPP stands for Strategic Lawsuite Against litigation filed by a large corporation or in a plaintiff to intimidate and silence a less power burdening them with the cost of a legal defense criticism. Not pretty.' Below the post, there is a section titled 'On Lisa Rein's Radar' with a sub-header 'First Amendment Under Attack' and a link to 'Things On Lisa Rein's Fair and Balanced Mind Today'. Further down, there is a section titled 'FORMER SPAM KING JERRY REYNOLDS (OF FARGO, ND) FILES SLAPP SUITS TRYING TO SQUASH THE FREE SPEECH OF PEOPLE EXPOSING HIS SPAMMY WAYS' with a date of 'November 07, 2005'. The text in this section discusses the 'No way, this can't be happening in America' department, mentioning David Ritz and Ed Falk being sued for exposing Jerry Reynolds as a spammer. It also mentions that you can download a zip of all the legal docs here and that Mary Hodder explains it. There are three small images showing legal documents. The text concludes with 'He's apparently filed (complaint here) SLAPP suits against two people, David Ritz and Ed Falk, who found that in the late 1990's, he was the largest spammer online (email wasn't so big then, but he had the largest porn spam operation on UUNET, with Sexzilla and Netzilla which were registered to Jerry Reynolds). And now he's using C&D's and these lawsuits to get whatever traces of the information that documents his spam and porn operation off of Google including search results and groups. He denies owning the site, btw, even though he was listed as administrative contact. Reynolds has even gone so far as to subpoena Ritz' and Falk's computers and put a gag order on one, but the other one is out of jurisdiction (the suit is in North Dakota, though Falk lives and does business in CA). Tomorrow, Ed Falk has to give a copy of all his computer harddrives to lawyers, who are still fighting over the jurisdiction issue. I think though that if this case went before a judge in CA, it would immediately be dismissed. All the evidence shows Reynolds was the'.

Jerry Reynolds, (former) Spam King (of Fargo,...
Jerry Reynolds, (former) Spam King (of Fargo, ND) Files Slapp Suits Trying to Squash Free Speech by People Exposing His Spammy Ways
IMG Jerry Reynolds, owner of Sexzilla.com in 1996 and 97 IMG
Jerry Reynolds, Owner of Netzilla in 1996 IMG John Doe v. Ed Falk
IMG Sexzilla, the Spamking's



Posted 14 days ago in [Napsterization](#) 379 links
[Search this blog](#)
By [mary hodder](#)

Post Blogging: #1 hit on Google

The screenshot shows a Google search results page for the query "jerry reynolds". The browser window title is "jerry reynolds - Google Search". The address bar shows the URL "http://www.google.com/search?q=jerry+reynolds&start=0&s" and the search bar contains "jerry reynolds". The search results are displayed under the heading "Web" and show "Results 1 - 10 of about 3,830,000 for jerry reynolds. (0.19 seconds)".

The first result is titled "Napsterization" and describes Jerry Reynolds as a former Spam King. The second result is "Jerry Reynolds Coaching Record - Basketball-Reference.com", which includes a bio of Jerry Reynolds (Ice) and his statistics. The third result is "KINGS: Kings Broadcaster Bios", and the fourth is "KINGS: Jerry Reynolds-Jack of All Trades". The fifth result is "Jerry Reynolds, Owner of Netzilla in 1996 on Flickr - Photo Sharing!".

On the right side of the page, there is a "Sponsored Links" section with two links: "Springfield, MO Attorney" and "Find This Person".

Finding Influencers

- Read posts, comments and follow conversations
- Find communities
 - Self tagged bloggers in Technorati by topic
 - Sphere lists
- Looks for link and subscription counts of those who link to you
- High numbers are indications of popularity (not authority)
- Look for key word and tag references to photos, blog posts, urls and to you...

In the end... you have to hack the blogosphere yourself

- You must read the posts, look at photos, del.icio.us links and tags
 - If people can find it online, you must be concerned about it, no matter the perceived influence
- But there is hope for better metrics!

What's needed

(Blog search is in many ways at the state of static web search in 1997)

- More sophisticated interfaces
- Topic browsing for small groups
- Sophisticated weighting tools
- Static web search adjustments

Thank You!

Mary Hodder

mary@dabble.com

Napsterization.org/stories

Dabble.com/blog