

THE INVISIBLE INTERNET

What you typically search and find using Google or Yahoo is only a fraction of the information actually available online. But think about it...offices use intranets and internal sites to share budget information, usernames, staff lists and more. All of that "hidden information" lives on something called the Invisible Internet. This tipsheet, along with my seminar, is designed to help you access the hidden 'net. This tipsheet is updated regularly. For the most current version, please see http://www.mydigimedia.com/tipsheets/invisible_net.pdf.

Web 2.0 Searching:

Looking for information in blogs, discussion boards, podcasts, etc.

Google, Google, Google. It seems like every time there's buzz about search, the story has to do with Google. There's a lot more out there, and while yes, Google is playing with serious math porn back in its top-secret development labs, there are some newer, more interesting ways to search through all that cyber sludge.

How does this help you, the traditional journalist? For one thing, Lexis ain't coding Podcasts to drop into its databases. If you're working on a story about the Japanese shrimp economy -- and at one point a few years back, I was -- I can guarantee that some shlub devotes his lonely nights to researching and recording a Podcast about them. Would I quote him as a source? Not a chance. But I'd take that information as a lead and start tracing his steps to find my analysts and my experts. Below, you'll find a handful of search alternatives. Try them out and see what you get!

Database of Web 2.0 Search Engines: http://www.allthingsweb2.com/mtree/SEARCH_2.0/

For Fun: Snoop Search at <http://www.gizoogle.com>

Search the Blogosphere

- **Blogdigger (<http://www.blogdigger.com>)**
Blogdigger is a powerful search tool that digs only through blogs to retrieve information.

A reporter's tool: Some bloggers/ blog sites typically break news before local journalists even hear about an event. Examples include TechCrunch and Slashdot. Find the bloggers who cover your beat and read what they're saying.

- **Technorati (<http://www.technorati.com>)**
This site indexes and tracks who's linking to who throughout the blogosphere.

A reporter's tool: You can search by keyword, by broad topic, by blog name and by Technorati user. An easy way to develop reporting on a specific topic would be to simply follow links.

- **Everyzing (<http://www.everyzing.com/>)**
Originally called *Podzinger*, this site indexes video and audio content.

A reporter's tool: Search by keyword – results are returned with the sentence where the keyword occurred and the time on the audio or video track. Rather than listening through an entire 45 minute podcast of a Hillary Clinton speech on health care, you could "zing" it to find the exact phrase and surrounding context – and then play the recording from that point forward.

Search the Zeitgeist

- **Omgili (<http://www.omgili.com>)**

Search through what people are saying in forums and discussion boards.

A reporter's tool: What are people saying about this particular topic? Are they mentioning other names/ companies/ websites? Will they point you more towards what you're researching?

- **Pixsy (<http://www.pixsy.com>)**

Mega-search for images and videos on dozens of journalism/ news sources, now divided by category.

A reporter's tool: Need to find a video to accompany your online story and have only 10 minutes left before deadline? You may have an easier time searching for video files here than on YouTube.

- **StumbleUpon (<http://www.stumbleupon.com>)**

StumbleUpon is actually a toolbar that can be installed on any web browser and, based on a set of your preferences, delivers new web pages you may like.

A reporter's tool: Just as looking through the footnotes of SEC reports can produce interesting story ideas, so too can rummaging through user comments. See what other people have said about a particular website and if their comments are relevant to your story.

- **Wikirage (<http://www.wikirage.com>)**

This site tracks pages in Wikipedia that are currently receiving the most traffic and edits.

A reporter's tool: When a story breaks, it may be worthwhile to check out what users are contributing at Wikipedia to help inform your own reporting.

Search People

- **Spock (<http://www.spock.com>)**

Spock is now in public beta and delivers thorough results on people. It pulls content from other websites and allows users to enter their own information, wiki-style.

A reporter's tool: Because this is a wiki, people have the ability to enter and edit information about themselves. Additionally, others may edit information, too, so you may be more likely to find leads on people here than by using a simple Google search.

- **Pipl (<http://pipl.com/>)**

Pipl searches the deep web to find information hidden within databases and other areas that standard web crawlers can't or won't search.

A reporter's tool: This search engine aggregates databases for you, which can be a plus when searching for information on deadline.

Search People (cont.)

- **Wink (<http://www.wink.com>)**

Wink is a smart search tool that pulls information only from social network sites such as MySpace, LinkedIn, Bebo and Friendster.

A reporter's tool: This is an easy, efficient way to search through social networks, however you will need to enter both a name and a location.

- **ZoomInfo (<http://www.Zoominfo.com>)**

ZoomInfo offers two search strategies, free and paid.

A reporter's tool: The site is geared to recruiters, but it offers plenty of leads for reporters, too. You can search by person, by a person within a company, or just by a company.

Search Across Networks

- **Infopirate (<http://www.infopirate.com>)**

Infopirate allows users to share their bookmarks.

A reporter's tool: Use bookmarks to look up information on individuals or companies.

- **Sputtr (<http://www.sputtr.com>)**

Sputtr offers single search window with many different options: Type in "Second Life" and search through YouTube, Flickr, Digg, etc. by selecting your button of choice.

A reporter's tool: Use this to find multimedia files very quickly.

- **Whonu (<http://www.whonu.com>)**

At Whonu, search through images, videos, news, maps, blogs, books, calendars, notebooks and more. After, you can share or bookmark your search results.

A reporter's tool: If your newsroom doesn't have an intranet, you can potentially use Whonu as a way to share reporting resources on a particular story, on a beat or on a geographic area.

More Search Ideas

- **Midomi (<http://www.midomi.com>)**

Need to find the name, title and artist of a song fast – but only remember the tune? Midomi allows you to hum a few bars of a song into your computer and it matches your rendition to its database, returning the original song information.

- **Gahooyoogle (<http://www.gahooyoogle.com>)**

Search both Google and Yahoo! at the same time using a split screen. It can be interesting to use this as a learning tool to discover the differences in how search engines return results. The creators of Gahooyoogle also recently launched **PolyCola (<http://www.polycola.com>)**, which offers the split-screen view with different search engine combinations.

Basic Web 1.0 Searching:

Looking for standard information in databases and in search engines

A good, general site on all things related to web searching:

<http://www.mach9design.com/deep/deep1.html>

Databases of databases:

<http://infomine.ucr.edu> (scholarly)

<http://www.completeplanet.com> (deep web)

<http://www.searchsystems.net> *NOTE: This is now a paid site.

Operators:

Search Term

"Amy Webb"
 football –eagles
 football ~glossary
 deli*

Returns...

The *exact* phrase Amy Webb with no variants
 pages with the word *football* but not *eagles*
 pages of glossaries about *football*
 pages with words that begin with *deli*

(Note: the asterisk * works better on some databases than others. Using * may or may not return good results searching Google or Yahoo!)

Google Greatness:

Search

City1 city2

#...#

book

define:

rphonebook:

phonebook:

movie:

stocks:

weather:

site:www.domain.com

link:domain.com

location:

source:

Returns...

Airfare and booking options

[\(Baltimore Chicago\)](#)

Search within a range of numbers

[\(Bill Clinton 1992...1993 speeches\)](#)

Search full text of books

[\(book Made to Stick\)](#)

Get a word's definition instantly

[\(define:optimize\)](#)

Get a residential phone listing

[\(rphonebook: Amy Webb Baltimore\)](#)

Get all phone numbers within a certain area

[\(phonebook:Chicago IL\)](#)

Get movie showtimes

[\(movie:300\)](#)

Get stock info

[\(stocks:MCD\)](#)

Get weather for a location

[\(weather:Baltimore\)](#)

searches within only the listed site

[\(site:www.irs.gov form 1040\)](#)

finds all pages linked to this domain

[\(link:www.mydigimedia.com\)](#)

get Google News articles from a certain location

[\(hot dogs location:Chicago\)](#)

get news articles from a certain source

[\(George Bush source:Washington Post\)](#)

Looking for difficult file types on Google:

Search

filetype:doc "fill in your query here"
filetype:ppt
filetype:pdf
filetype:xls
filetype:mbd

Returns...

documents
power points
PDF files
Excel Spreadsheets
Access databases

Title Searching: when looking for reports or publications

On Google use the command **intitle:**

URL Search: when looking within a web address

Google uses **inurl:**

Domain Searching: when looking for information within a certain kind of site

Usually it's **site:gov**

Common domains: gov (government), mil (military), edu (accredited colleges and universities), org (no meaning), co.jp (Japan), co.uk (UK), etc.

To look up the owner of a domain, go to the WHOIS feature on <http://www.networksolutions.com>

Finding missing websites/ pages:

<http://www.archive.org/>

Advanced Google Tips:

- **allinlinks:**www.xxxxxx.com = Searches only within links, not text or title.
- **allintext:**www.xxxxxx.com = Searches within the text of pages, not in links or titles

Put it together!

EXAMPLE: I want to get a copy of Jones University's budget, which isn't available from any of its websites.

SOLUTION: In Google, search:

site:www.jonesuniversity.edu filetype:xls budget fy 2007

You can also try adding:

site:www.jonesuniversity.edu filetype:xls budget **inurl:admin**

For even more Google tips and hints, have a look at the Google Advanced Search site:

<http://www.google.com/help/refinerearch.html>