

Create, Aggregate, Filter, Spy & Scrape

Prepared By Tracy Mikulec of Immersion Media
February 7, 2009

Tools and techniques to scour the Web for content and news, track competitors, and monitor social buzz, industry blogs and company Websites.

This guide gives content and Web editors a strategic advantage with:

- tools and techniques for targeting a specific niche
- ways to efficiently identify breaking stories
- secrets for discovering essential information before competitors do
- methods for finding interesting and compelling content that can be reported on, featured or aggregated on their Websites

The information dashboard: An editor's best friend

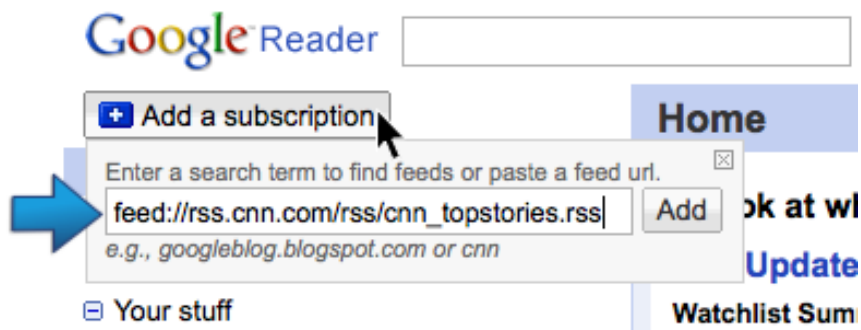
Typically, most content editors have a handful of blogs and sources that they scan on a daily basis, but going to each site can be time consuming. Creating a dashboard that conveniently aggregates all relevant blogs, websites, social media sites, company sites and that also performs custom searches on important industry leaders, athletes, events, companies and competitors is a very powerful and efficient way to monitor everything relevant in your niche and can create a major competitive edge.

There are many ways you can create an information dashboard, but Google Reader and iGoogle are the best places to start. Both collect RSS feeds from all the blogs, Websites, and searches you're interested in and will give you a home to view and organize all this information.

What is RSS? - "RSS in Plain English."
commoncraft.com/rss_plain_english

This document focuses on Google Reader, so go ahead and set up an account at: <http://google.com/reader>. Because this will be your intelligence gateway, you'll likely want to make this your home page.

Adding sites to your dashboard - Now that you've created your dashboard, you need to get information into it. First you'll want to add the sites you're currently following. Look for and click on the RSS or XML text/icon on a site. When clicking on the RSS link, each browser will handle it differently, but the ultimate goal is to add that RSS link to your Google Reader. If your browser doesn't ask you to add to Google Reader, copy the RSS URL (example: <feed://rss.cnn.com/rss/cnn.rss>) and paste it into your Google Reader.



Example of adding an RSS feed to Google Reader

Site doesn't have RSS feed? - If the site you're interested in doesn't have an RSS feed or RSS icon, feedwhip.com or page2rss.com services can create an RSS feed to add to Google Reader.

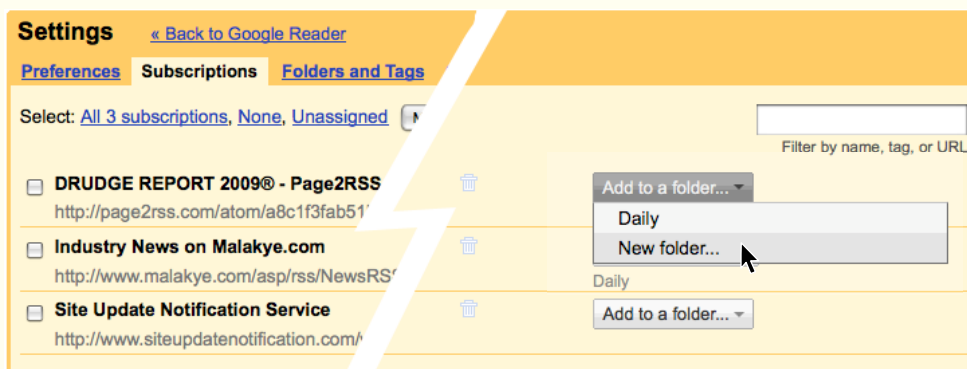
Folders - Keeping it organized

Google Reader allows you to create folders, so you'll be able to designate which folder you drop each new feed into and keep your information organized.

- There are many different ways you may want to organize your content, but to start, set up at least two new folders. Label one folder "Daily" - this one is for your top 10 to 20 sources of content... those that have the most relevant content in your industry or niche.
- Label the second folder "Archive." This contains all of the other blogs, RSS feeds and content that you want to monitor. Make sure to assign each blog and RSS

feed to one of these two folders. And remember, you can always add to and change these designations as your list evolves.

Create a Folder by clicking “Manage subscriptions »” (lower left hand corner) and selecting “New folder...” from drop-down (see illustration below)



Strategic tools for finding interesting content

Now you'll want to uncover additional blogs, Websites and social media sites that are specifically relevant or interesting. The challenge is to find and follow enough sites to avoid missing something important, yet follow a small enough number so as not to get overwhelmed. Here are some tools to really make you appear all seeing and all knowing.

Google News Alerts (<http://news.google.com/>) - Google News Alerts will search all news media and AP articles. Have results sent to you via RSS instead of in email. Create multiple alerts and put popular keywords alerts in your daily folder and all others should be put in the archive folder.

Alltop (<http://alltop.com/>): A Website that aggregates top blog posts, organized by topic. Just find your topic, and start exploring new sites. If you find a great site, add the RSS feed to Google Reader.

Google Blog Search (<http://blogsearch.google.com/>): Use Google Blog Search to capture the conversations happening in the blogosphere. If Google Blog Search yields too many splog results (spam blogs automatically generated by keyword searches), try using Twingly (<http://www.twingly.com/>), Icerocket (<http://icerocket.com/>), or Yacktrack (<http://yacktrack.com/>) as alternative blog search engines. They seem to be more spam free

Monitor the Social Buzz

An entirely new source of information—so new, in fact, that most editors haven't even realized it yet—is content and conversation from social networks. Much of it will be noise and not applicable, but once you learn to watch for patterns, social buzz can tip you off to controversy about your brand, an industry manufacturer or your competitors' sites. Post articles about what people are talking about, what's hot, or controversial. The discussion and buzz become the story.

Discussion boards: use Boardtracker (<http://boardtracker.com/>) or Omgili (<http://www.omgili.com/>) to conduct a search of topic-specific conversations across several different forums. After searching, subscribe to the results via RSS. You'll likely want to put the resulting feeds in the archive folder.

Yahoo Pipes. Yahoo Pipe Searches are powerful cloud applications that can search multiple sites for the keywords that are relevant to you, and post the results to your Google Reader. Searching can take a minute or two, but once the results are displayed, click the button to add the keyword search to your Google Reader.

Social Media Firehose (http://pipes.yahoo.com/update_maker/social_media_fire_hose): This tool creates one RSS feed that aggregates results from Flickr, Digg, YouTube, FriendFeed and other social media sites.

Latest Blog Mentions Pipe (http://pipes.yahoo.com/pipes/pipe.info?_id=bFH0.Di32xGnPkr_qu5lkA): This is another Yahoo Pipe that will aggregate brand references across several major blog search engines, including Technorati, Icerocket and Google Blog Search.

Twitter Search (<http://search.twitter.com/>): Twitter is a social network where people chat about what's hot around the Web (among other things). Monitoring its search function can give you a good idea of what's happening around a given keyword. Just type in your search term (industry buzzwords, your brand name, or your competitors), conduct the search (making sure to limit your language parameters as appropriate), and then subscribe to the RSS feed for the results. Place the keyword results in your daily folder and archive the other searches.

Set up a spy network for pages like press releases, company info pages and employment sections - Many company sites will have pages that aren't frequently updated, but you'll be interested in knowing when and what they update. This can tip you off to a new hiring position (the implications can mean they're firing someone or opening a new department), early knowledge of a press release or RFP announcement. Followthatpage.com and siteupdatenotification.com/watch will track sites, identify pages and content that has changed and give that information to you via RSS feed that you can include in your Google Reader.

Translate Foreign Sites - You may find interesting or related sites in another language or want to follow content in another part of the world. Get the RSS feed from the site and run it through Mloovi.com. This will create an English translated RSS feed to put into Google Reader.

Search Strategy

You should approach the topic of automated search with an overall strategy that is deliberate and focused. The strategy should take into account all of your known informational needs and prioritize them around searchable categories. This master search list will become your roadmap to long-term success. As the work of developing search strategies for each category progresses, additions and deletions to the master list can be weighed against others for overall importance.

It is important that you develop your search strategy before you ever go online. This means forming good search questions and keywords on paper (or other form) without the distractions of the online environment. This offline work provides a fixed checkpoint by which progress and search efficiency can be measured without the trappings and distractions of the online search query building process.

For a successful automated search, you must choose appropriate keywords and queries. That means putting sufficient research into finding keywords and spending time building, querying, testing and measuring the results you get. The goal is to make these search queries return fewer but more relevant results. It will take some time before you're really good at it, but once you find exactly which keywords yield the results you're looking for, you'll discover a host of blogs, Twitter profiles and videos directly relevant to your topic area. Appendix A provides specific information, examples, strategies, and resources for developing your overall search acumen.

Remember that building your content sources is a difficult process that takes time to do well. Be patient.

Review your dashboard daily

Plan on spending 30 minutes a day skimming and reading your daily Google Reader folder. Use the “J” and “K” keys to quickly cycle back and forth between posts (for a list of other Google Reader keyboard commands, see Appendix C). Click to read more on the posts that interest you.

Next, go to the search bar at the top of the Reader window. Conduct a search for the keyword(s) you've found to be effective for you. This helps you avoid searching the entire blogosphere – rather, it limits the search to the content you've already deemed important (those that have already made your RSS Reader). Using the Google Reader keyboard shortcuts, you should be able to cycle through and skim these posts quickly, stopping only to read the most relevant items (things you think you can use).

Share with your editorial team

Empower your editorial team with this information, and set each of them up with a Google Reader account. Assign each editor a specialty area, and help them customize their account with keywords specific to their area of interest. For example, an editor that focuses on product reviews will have different keywords, sites and searches than an editor specializing in event coverage.

You can also export your list of blogs (<http://www.google.com/help/reader/faq.html> called an OPML file), which will help them to get started quickly, and gives them an opportunity to hone their results as they go.

Conclusion

By spending 20-40 minutes a day monitoring relevant Websites, blogs, and content searches, you'll be positioned to identify, aggregate and present the most comprehensive snapshot of everything interesting in your industry. And of course, it means you'll be well informed and equipped to apply all this knowledge toward your in-the-trenches journalism, investigative reporting or hard core corporate intelligence.

Sources: Mashable.com, Google.com, web-strategist.com, readwriteweb.com

Appendix A

Developing Search Skills and Strategies

You should get started using the search tools outlined above (Google Reader, GoogleNews, Yahoo Pipes, etc) without worrying if your search skills need to be improved. Google and Yahoo search functions are simple to use, and what's more, they offer very good results without knowing any advanced search techniques. For example, Google is so good that if you type just a few targeted keywords, you'll be able to fish through four or five pages of returned results and find what you're looking for.

However, you will soon find that investing some time in developing your approach to search (your "search strategy") and learning more advanced search functions will pay huge dividends in finding great content. It will allow you to shorten your time spent searching, shorten the amount of time needed to set up your search tools and, most importantly, it will limit your results to a small number of incredibly accurate results. Again, basic search will get you pretty far, so don't delay getting started.

Making Your Search Strategic

First, it's highly advisable to start your search strategy offline. Of course, you'll need to go online to test some of your strategies, but having an offline approach ensures that you stay focused, on track, and it will limit your frustration. Try creating a special folder or notebook to record your strategy and progress. The benefits will become obvious as you work through these steps. You can also record your strategies in other digital formats such as spreadsheet, databases or word documents, but they are not preferable because they don't offer the same barrier to "going online" because your browser is a few keystrokes away.

Before You Begin

First, make a list of all the possible topic areas you want to research. Then prioritize this list (you only have so much time and you want to be as effective as possible). This list will become your road map as you begin searching and finding new things to research or your priorities shift.

Examples:

Surf trips to Australia	Sports teams	Lyrics
Skate videos	Video	X Games
Competitors	Photography	Olympics
Advertisers	Writing	Poetry
Interesting people	Music	New truck

Step 1: Create a “research question”

Choose a research area of focus and state what you are looking for in the form of a question with as many specifics as possible. If your search topic is more complex, start with a series of questions.

For example, using the “new truck” topic from above:

What car should I buy if I want an economical, fuel-efficient truck that can off-road and carry a 3,000-lb. payload?

Step 2: Develop keywords

Underline keywords from your search question(s) above.

What car should I buy if I want an economical, fuel efficient, truck that can off-road, and carry a 3,000-lb. payload?

Step 3: Broaden your own keywords

Using the keywords that you've outlined above, flesh out others using your own knowledge and experience. Consider synonyms and other similar words.

For example:

For “buy” you might add “purchase” and “lease”

For “economical” add “price”

For “fuel efficient” add “MPG”

For “truck” add “SUV”

For “off-road” add “4-wheel drives”

Also jot down sources that you are already familiar with, such as Consumer Reports, Moody's, Car and Driver etc.

Step 4: Go online to add more and better keywords

Do some quick "Google passes," entering some of these keywords into Google's basic search engine and looking for keywords or terms you haven't considered. Use the sources you outlined above. (This is where keeping your search notes offline is extremely helpful. You will be tempted to jump right into your search--don't do it.) These search passes will help you pick up new keyword ideas you may not have thought of before.

Another helpful hint for finding keywords is if you find a site that gives you good information during your "Google passes," simply use the "similar pages" link at the bottom of the search result.

Another hint is to do a simple keyword search using Google's adsense keyword finder site, which yields more than 200 keywords related to truck, SUV, and MPG <https://adwords.google.com/select/KeywordToolExternal>

This extra effort should yield additional keywords. In our example, we were able to find "crash test rating," "MSRP," and "4WD." Many times an industry or topic area will have specific terminology that will be very helpful in narrowing down your search results. Add those terms to your keywords.

Step 5: Keep File Formats in Mind

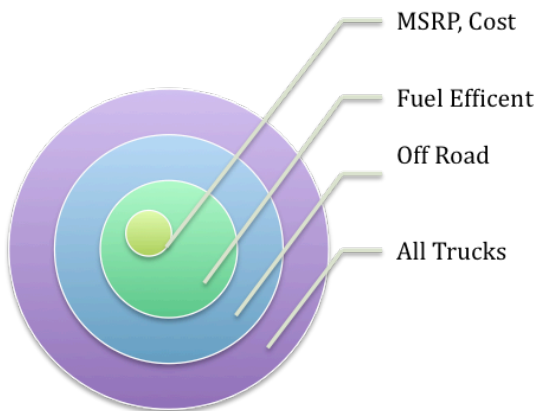
If your particular search strategy involves specific file types, such as videos, you can use the file extension as one of your keyword search terms. These can be helpful for finding videos, news broadcasts, podcasts, lectures, photos and images. This is particularly helpful if you are developing a feed that is only for rich media, newly posted videos, or podcasts, etc. for one of your search categories.

Media File Type	extension	To Find
Audio	.MP3	Songs, podcasts, lectures
Movies	.MOV MPEG	Videos, news broadcasts
Documents	.DOC and .PDF	Research, reports, white papers
Presentations	.PPT	Presentations, meeting notes
Rich Media	.SWF	Animations, Video
Image	.JPG	Photos and Images

Step 6: Create logical search order

The order of in which words are put into search engines is very important. Organize your keywords into logical sequence. This hierarchal approach lists keywords from broader topical areas first down to narrower terms last. (Sketching out the keywords using circle diagram is useful exercise.)

For example: "Trucks" is a broader search term than "off-road" (off-road trucks are just one of many types of other trucks), then there's expensive and less expensive off-road trucks within that grouping and so on. "MPG" and "cost efficient" have to come after those because they are not the primary focus of the search.



Google puts more emphasis on the words that are first in the search order. Therefore, the hierarchal listing in this example (trucks off-road fuel efficient MSRP cost) will give you results that are more focused on trucks and off-roading than you would get pages focused primarily on pricing, fuel efficiency and off-roading trucks in general.

Another slight variation on this strategy is to list the keywords in the order that you expect that they would appear on the types of pages you are looking for.

So for example, if we know for sure that we are looking for off-road enthusiast sites, we might use "off-road" or "4WD," rather than "truck" as the first keyword. But while this approach is many times more useful it is sometimes better to start with the hierarchal approach as a first step then refine the position of a few of the keywords.

Use a matrix such as the diagram below to keep track of the keywords that you are searching, replacement terms, and the order. Keep these search strategies in your workbook folder.

Keyword 1	Keyword 2	Keyword 3	Keyword 4	Keyword 5
Truck	New	Off-road	Economical	Fuel Efficient
SUV	Used	4X4	Value	MPG
	Pre-owned	Four by four	Entry level	
		4WD	Low cost	

Try various combinations of your search terms and record the sequence in your matrix. Google allows you to search up to 32 keywords. However, the more keywords the broader the search. Keep it short and well defined.

Step 6: Trial and Error, Rate your Queries.

Refine your queries through trial and error.

Example, using the search query:

Truck, 4X4, pre-owned, fuel efficient, MPG

You may find that the keyword "pre-owned" in the query is returning a lot of car dealerships. That just isn't working out, so note that in your query matrix.

Or you might find that the keyword "4WD" returns better sites than using the term "4X4", so you note that.

As you drop certain search terms in your search query matrix use a rating system to indicate the queries' usefulness. Use a simple 4 or 5 star rating system.

Keyword 1	Keyword 2	Keyword 3	Keyword 4	Keyword 5	Rating	
Truck	New	Off-road	Economical	Fuel Efficient	•••	
SUV	Used	4X4	Value	MPG	••	
	Pre-owned	Four by four	Entry level		•	
Truck		4WD	Low cost	Fuel Efficient	••••	

Tracking and rating your queries will become important because you will forget which search terms you used, and how successful they were, especially if you put down your work and are not able to get back to it immediately. (Once searches are set up to aggregate content in your feeder, you may not get back to them for weeks or longer.)

Step 7: Refining Search and Advance Operations

Google has a series of special operators that help narrow your search results even further. The three essential ones are quote marks (""), the plus symbol (+) and the negative symbol (-).

Using our example above, we can use the "" to return pages with a specific phrase (Truck, 4WD, "fuel efficient"). Without the parenthesis, Google searches for the word "fuel" separately from efficient.

To make sure any Web page we get definitely has the word "MSRP," we can simply add +MSRP to the search term. Google will return only pages that have that term.

If we want to make sure we don't get certain truck manufacturers, we use the negative sign to remove them from the results, such as:

truck, 4WD, "entry level" -Ford -Nissan

Again, add these keyword limiters to your search query rating system so you can evaluate their effectiveness.

Step 8: Label, Monitor and Refine

Make sure to label your searches with a descriptive title. Keep track of these queries along with your master search strategy mentioned at the top of the page.

Keep a list of all the searches you want to eventually build. This is important because you will get distracted as you go online and test your query results and build keyword lists. At some point you may want to abandon your search topic altogether and take it in a whole new direction or explore new ideas. Resist the

urge and simply write them down for safekeeping (also bookmark sites that you find interesting during your search). Consider changes to your search strategy offline and in your prioritized global search strategy. Monitor the amount of time that you are spending on any given topic. Question yourself, "Is this worth the time I've invested? Am I getting the results that I want?"

Move these searches into a folder labeled "BETA feeds" or "new feeds" in your feeder. You may also want to name the file itself as "Trucks Beta." This is a reminder that these feeds are being monitored (by you) to see how effective they are. Once they are working properly giving you good data and not cluttering up your feeds with useless false positive results, then move it into your normal folders to check.

Once you have your queries set up and data is being fed into your feeder, you may find that you are not getting the results you expected. A simple approach is to go back to your matrix and start substituting new keywords. Go back, revise your rating system, and look for approaches you may have missed. (This is when keeping good offline records becomes crucial.)

If those attempts are still not getting what you want, you may need to get into more advanced searching operators. Different search engines have different advanced operators, which are listed in each engine's Advanced Search page.

Some types of advanced operators include wildcards, look in URL, look in the title, limit search to types of files, date searching and, for many sites, keywords that must be near other keywords on the page.

Online Search course:

<http://movielibrary.lynda.com/html/modPage.asp?ID=108>

Link to: Google's Search help

<http://www.google.com/support/websearch/bin/static.py?page=searchguides.html&ctx=basics>

Google Operators cheat sheet for other advance search operators

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

Appendix B

Create a Reference Search Engine

www.google.com/coop/cse/

Create an incredibly powerful custom search tool that searches only the sites you define. Using Google Custom Search, limit your research to a specific category by listing all of your top sources in your field. Set up by going to google.com/coop/cse/, and adding the URL's, save it, then bookmark it. Now any time you want to research or find information among your top sites, you can just search for keywords with your Custom Search Engine.



Create a Custom Search Engine

1. Set up your search engine
2. Try it out

Basic information

Give your search engine a name and provide a brief description.

Search engine name:

for example, Real Climate Search Engine

Search engine description:

for example, Climate Science from Climate Scientists

Specify some keywords that describe the content or subject of your search engine. These keywords are used to tune your

Search engine keywords:

for example, climate "global warming" "greenhouse gases"

Search engine language:

What do you want to search?

- Only sites I select.
- The entire web, but emphasize sites I select.
- The entire web.

Select some sites

Specify a list of websites to search. You'll be able to edit this list and add more sites later. [Tips on formatting URLs.](#)

Sites to search:

List one URL per line

Appendix C

Google Reader keyboard shortcuts

Keyboard shortcuts save you time by helping you quickly explore your reading list without moving your hand back and forth between your keyboard and mouse. Here are the keyboard shortcuts for Google Reader:

Keystroke	Action	Definition
j	next	view the next item in your reading list
k	prev	view the previous item
n	scan down	scroll to the next item without viewing it
p	scan up	scroll to the previous item without viewing it
shift n	page down	scroll down by one screen of items
shift p	page up	scroll up by one screen of items
o or <Enter>	view item	view the selected item
h	top	go to the top of your reading list
r	refresh	refresh your reading list
s	star	star the current item
l	label	jump to the form for labeling the current item
v	show original	open original item in a new window
g then h	home	loads your home (reading list) items
g then s	starred	loads your starred items
g then u	edit subscriptions	loads your subscriptions and labels
g then r	read items	loads your read items
g then l	label selector	lets you type in a label and then loads it



34571 Calle Rosita, Capistrano Beach, California 92624
Phone 949-422-4875 | www.immersionmedia.com

©2009 Immersion Media. All Rights Reserved.