YAHOO! HACKS™

Tips & Tools for Living on the Web Frontier

O'REILLY®

Paul Bausch

HACK #88 Visualize News Topics as Tags

Two services, Yahoo! News Tag Soup and TagCloud, use the Yahoo! Contextual Web Search API to help you visualize which topics are mentioned the most across Yahoo! News and your favorite sites.

As you consume the news each day, you get a general sense of what the big stories are. If CNN is leading the hour with news about the Supreme Court, you open the newspaper to find a story about the Supreme Court on the front page, and then you spot a Supreme Court story at the top of Yahoo! News, you can bet the Supreme Court is the hot topic. But this general way of identifying a hot topic isn't as precise as actually analyzing the content itself to quantify which keywords are mentioned most often throughout a day.

With digital information, it's possible to track exactly which words are used more frequently than others. And there are a number of ways you can categorize information to make that tracking easier. One system of categorization is called *tagging*, in which people add keywords to an article, photo, or web site in order to organize that bit of information. Each keyword is referred to as a *tag*, and lots of fun can be had analyzing the tags.

A popular way to visualize tag usage is in the form of a *tag map* that displays more popular tags in a larger font. So popular tags appear large, and less-popular tags appear small. With one glance at a tag map, you can see which topics are hot.

Now, imagine you have a source of information, but no people to tag it for you. That's where the Yahoo! Contextual Web Service can help. This service lets you send an arbitrary amount of text, and it will automatically extract keywords from that text. The service ignores common words such as *a*, *and*, or *the* and returns the more unique words from a text. For example, the Yahoo! Contextual Web Service analysis of the text of this hack returned phrases such as *supreme court*, *hot topic*, and *search api*.

By automatically extracting keywords from text, you can create computer-generated tags and perform similar sorts of analysis on those tags.

Yahoo! News Tag Soup

Seeing a more precise picture of news keywords was the motivation behind a service by John Herren called Yahoo! News Tag Soup (*http://yahoo.theherrens.com*). By sending stories from Yahoo! News through the Yahoo! Contextual Web Service and storing the resulting keywords, you can visualize trends in the day's news as a tag map.

Figure 5-25 shows the Yahoo! News Tag Soup for June 28, 2005.

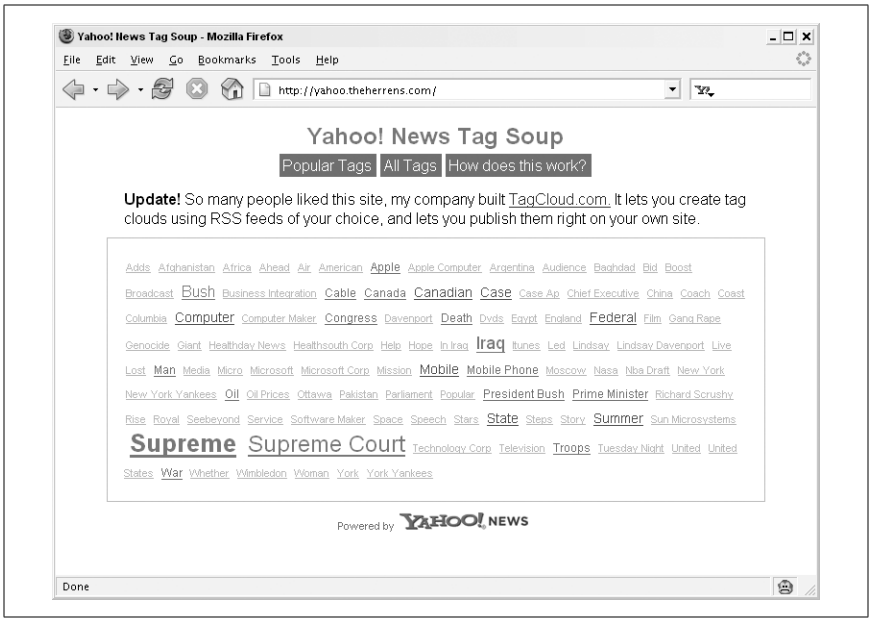

Figure 5-25. A tag map with topics culled from Yahoo! News stories

As you can see, the Supreme Court was in the news that day. When you click on the tags at Yahoo! News Tag Soup, the service provides a list of stories that contain that particular keyword, as shown in Figure 5-26.

Yahoo! News organizes stories by importance, or into broad categories such as politics, business, or technology. Yahoo! News Tag Soup is a completely new way to browse news stories, made possible thanks to Yahoo!'s open technology.

Not content to focus solely on Yahoo! News, the creator of Tag Soup helped build a way for you to assemble your own personal news visualizations from any source.

Your Personal Tag Cloud

John Herren's company IonZoft, Inc., took the idea behind Yahoo! News Tag Soup and made it accessible to others. The application is called TagCloud (*http://www.tagcloud.com*) and it's a free service. To get started, browse to the TagCloud web site and create an account.

Once you're logged in, you can create a personalized cloud. The Cloud Name must be unique across the system, so you might need to choose something unusual. For instance, the cloud name *oregon* is probably taken by someone else already, but *oregon2112* is probably available. Use the Description to

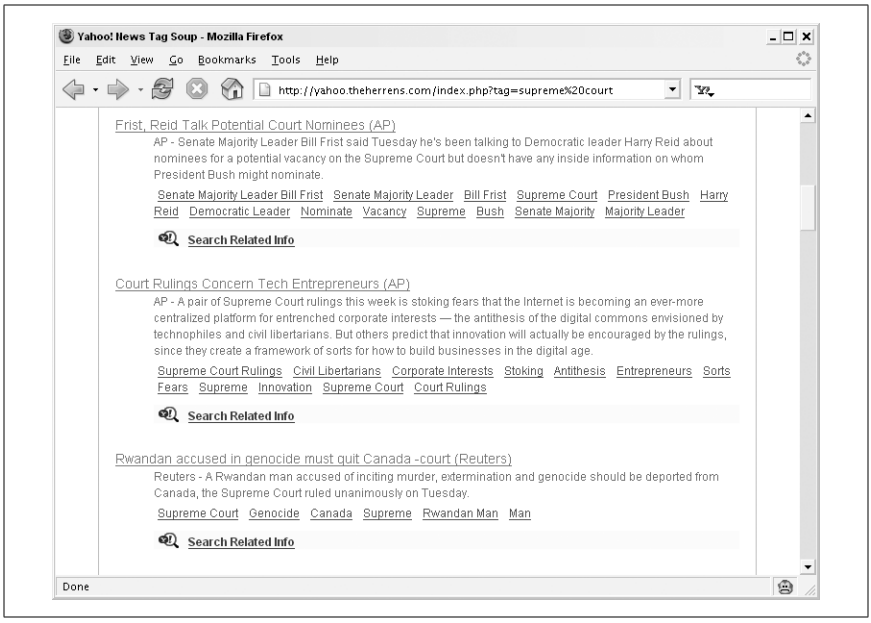

Figure 5-26. A list of Yahoo! News stories by tag

identify what types of news sources you're going to analyze with your cloud. Click Create Cloud and you'll have an empty space you can start filling up with personal news and information.

TagCloud relies on RSS feeds as input, and you can add individual feeds one at a time by clicking the Feeds tab and entering the URL. If you already have a list of feeds you read at a service such as Bloglines (*http://www.bloglines.com*), you can import several at once by clicking the Import OPML tab and specifying an OPML file that contains a list of your feeds.

Once you have a few feeds included in your cloud, click the View tab at the top of the page. You should see a list of tags that have been extracted from your feeds, with each tag a different size based on its frequency, as shown in Figure 5-27.

Click on any tag to see a list of matching stories and posts across your feeds. This is a useful way to visualize what topics are hot on a specific group of sites. You can also publish your TagCloud to your web site, so you can share your slice of the news with others.

Click Implementation Guide at the top of any TagCloud page and you'll see how easy it is to add a TagCloud to a remote site. You simply need to copy two lines of JavaScript and paste them into your site. The TagCloud service

Figure 5-27. A personal TagCloud

does the rest. Figure 5-28 shows a personal TagCloud included as a sidebar on a weblog.

As readers click on a tag, they'll find a list of posts and stories from your selected feeds that have that keyword.

Listing your personal TagCloud on your weblog, or reviewing it privately for fun, is a great way to keep a visual handle on the day's news.

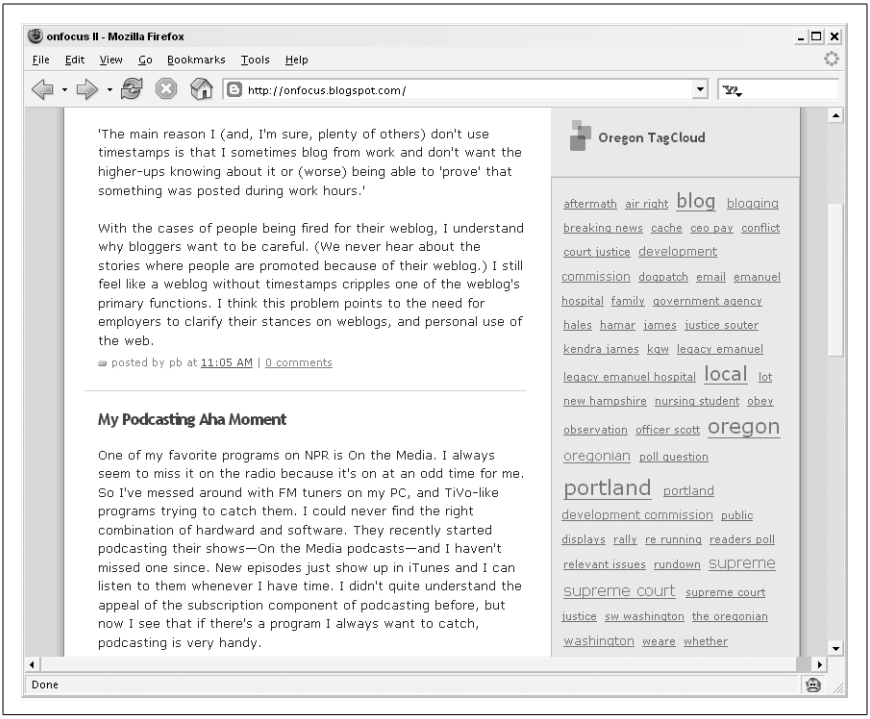

Figure 5-28. A personal TagCloud on a weblog