

What's Up with Tag Clouds?

Marti A. Hearst

Visual Business Intelligence Newsletter

May 2008

If you've read blogs, visited social bookmarking web sites, or used a photo sharing web site such as flickr, you have probably noticed a rather eye-catching visualization in which words and phrases of differing sizes float in space. This representation is usually called a *tag cloud* (but is also known as a *weighted list*). I show a few examples below:

accessories apparel art
arts audio australia baby
bags beauty books
business california
canada car casual cheap
children clothes
clothing community
companies computer
computers
computersandinternet
consumerelectronics

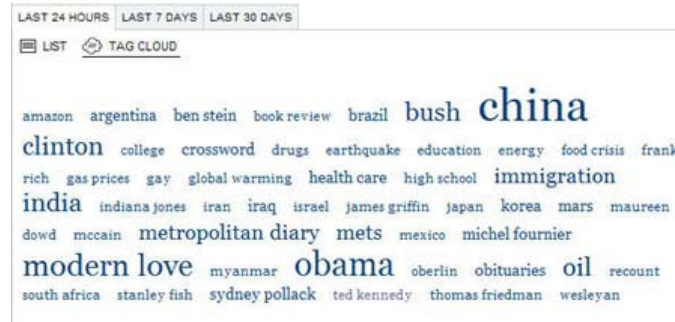
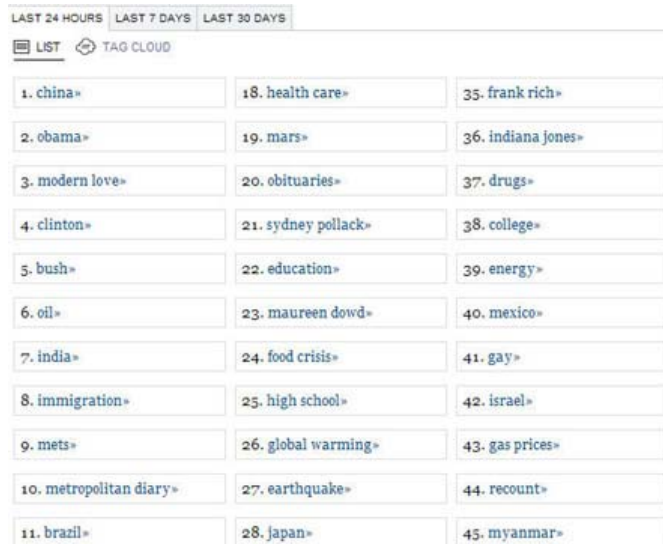
3d accessibility actionscript advertising ajax art
blogger branding business cms consultant creative
css-design design designer developer deve
drupal e-commerce ecommerce flash freelan
freelancer graphic graphic-design graphics hc
identity illustration illustrator internet java javascript
logo logo-design marketing mysql photographer pho
photoshop php php-mysql-developer portfolio print

rica aided ally anbar armed army baghdad bless challenges
nders commitment confident confront congressman constitution
cratic deploy dikembe diplomacy disruptions earmarks economy
mists failing faithful families freedom fuel funding god haven
iraq islam julie lebanon love madam marine math medicare
lestinian payroll province pursuing qaeda radical regimes resolve
te september shia stays strength students succeed sunni tax

(Images from www.SmashingMagazine.com)

The pieces of text are known as *tags*, which are labels attached to images, web pages and blog postings. These labels are usually quite short, and often consist of a single word or a pair of words stitched together with an underline or no separator at all. Sometimes tags are referred to as “social tags”, because the viewers of the information objects are often the people who assign the tags to those objects. People assign tags to information for a variety of reasons. In the case of their own information collections, tags are assigned to organize the information and make it easier to find (and this is especially important for photographs which do not have textual metadata). But in the case of assigning tags to *other people's* information, as is commonly done on social media websites, people sometimes tag “in order to communicate contextual information to others about the image and consequently about themselves” (Ames & Naaman 2007).

When they first started appearing, tag clouds really puzzled me. Perhaps given my background teaching information visualization and data graphics, my initial thoughts were that tag clouds were an attempt to use visualization to provide insight about the tags that cannot be understood from a simple list of words. But there are simpler, clearer ways to show this information, such as a simple list of words in frequency order. The website of The New York Times shows most frequently searched on words as either a list or a tag cloud:



(Images from www.NYTimes.com)

I was confused about tag clouds in part because they are clearly problematic from a perceptual cognition point of view. For one thing, there is no visual flow to the layout. Graphic designers, as well as painters of landscapes, know that a good visual design guides the eye through the work, providing an intuitive starting point and visual cues that gently suggest a visual path.

By contrast, with tag clouds, the eye zigs and zags across the view, coming to rest on a large tag, flitting away again in an erratic direction until it finds another large tag, with perhaps a quick glance at a medium-sized tag along the way. Small tags are little more than annoying speed bumps along the path.

In most visualizations, physical proximity is an important visual cue to indicate meaningful relationships. But in a tag cloud, tags that are semantically similar do not necessarily occur near one another, because the tags are organized in alphabetical order. Furthermore, if the paragraph is resized, then the locations of tags re-arrange. If tag A was above B initially, after resizing, they might end up on the same line but far apart.

Tag clouds also make it difficult to see which topics appear in a set of tags. For example, in the image below, it's hard to see which operating systems are talked about versus which ones are omitted. Intuitively, to me, it seemed that an ordinary word list would be better for getting the gist of a set of tags because it would be more readable.



(Image from del.icio.us)

Only two scientific studies have been done comparing tag clouds to more standard lists, and in both these studies, tag clouds cast a shadow over the information, in comparison with simple lists (Rivadeneira et al. 2007, Halvey & Keane 2007).

But despite their perceptual flaws, tag clouds are appearing all over the web, and I wanted to understand why. I decided a good way to understand what tag clouds are for would be to ask people whose websites use tag clouds. At a meeting in 2006, I interviewed 15 Web 2.0 leaders and experts, and after this I conducted a somewhat systematic analysis of what people say when they write about tag clouds on the web (Hearst & Rosner 2008).

The main insight I gained from the interviews and analyses was the following: tag clouds are meant to show that there are *people actively using* the information, commenting on it, and categorizing its contents. They indicate mental activity and social activity, and their visual appearance reflects this. Their irregular shapes, the white space between the words, suggests human movement and spacing, like people of different shapes and sizes standing around a room chatting at a cocktail party. They are meant to signal that people are there, and they are talking. The tags themselves give perhaps just a glimmer of what the people are talking about, like snatches of conversation overheard at the party.

I also found that people thought that clouds built by editors or machine-generated tags were not “valid”. One blogger chastized a parenting web site whose editors themselves had created tags representing their products, denouncing the faking of user activity.

Once you gaze at tag clouds from the social point of view, they make sense. Although some commentators claim they are more useful than a plain list for navigating lists of tags and understanding the gist of the site, this claim is not supported by the empirical evidence. But if you see them as a signaller of social mental activity, then they work quite well.

It turns out that an early observer and promoter of this idea was a researcher at the MIT Media Lab named Judith Donath. She and her students have been studying the visualization of social interaction since the mid-nineties. In a paper published in 2002, she noted explicitly that:

Traditional data visualization focuses on making abstract numbers and relationships into concrete, spatialized images; the goal is to highlight important patterns while also representing the data accurately. This is a fine approach for social scientists studying the dynamics of online interactions. Yet for our purpose it is also important that the visualization evoke an appropriate intuitive response representing the feel of the conversation as well as depicting its dynamics. (Donath 2002)

Donath describes a conversation visualization tool developed by her student Dana Spiegel which “creates an environment that looks like the conversation.” Although tag clouds do not closely resemble the researchers’ proposed visualizations, the spirit behind them is the same. And unlike the research visualizations, tag clouds are very easy to implement on a web page.

Something else that came out of my interviews and analyses is that some people think tag clouds are good at showing “trends” in the subject matter that they label. In information visualization, we often think about visualizing trends as depicting patterns of change over time. Tag clouds only show one snapshot in time, so this seems to be a contradiction at first glance. But there is another sense of trend which means “intrinsic” or “a general tendency or inclination,” and I think it is this sense of trend that people mean when they discuss tag clouds.

It’s an open question whether or not tag clouds will stay popular, and for how long. Many people see them as trendy and faddish, and some of the bloggers remarked that they think they are uncool now that corporate websites are using them. I suspect that they will stick around for a few more years, in large part because they are easy to implement and web crawlers can analyze the text. I also suspect that after a few years they will look dated and a different way of signaling social activity will become popular and replace them.

References:

- Ames, M & M. Naaman. "Why we tag: motivations for annotation in mobile and online media." CHI 2007.
- Donath, Judith S. "A semantic approach to visualizing online conversations." Commun. ACM 45(4): 45-49 (2002).
- Halvey, Martin and Mark T. Keane. "An assessment of tag presentation techniques." WWW 2007: 1313-1314.
- Hearst, Marti A. and Daniela K. Rosner. "Tag Clouds: Data Analysis Tool or Social Signaller?" HICSS 2008. <<http://people.ischool.berkeley.edu/~hearst/papers/tagclouds.pdf>>.
- Rivadeneira, W., Daniel M. Gruen, Michael J. Muller, and David R. Millen. "Getting our head in the clouds: toward evaluation studies of tagclouds." CHI 2007: 995-998.

About the Author

Marti A. Hearst is an associate professor in the School of Information at UC Berkeley, with an affiliate appointment in the Computer Science Division. Her primary research interests are user interfaces and visualization for search engines, computational linguistics, and empirical analysis of social media.

She received BA, MS, and PhD degrees in Computer Science from the University of California at Berkeley, and she was a Member of the Research Staff at Xerox PARC from 1994 to 1997. Prof. Hearst is on the editorial boards of ACM Transactions on the Web and ACM Transactions on Computer-Human Interaction and was formerly on the boards of Computational Linguistics, ACM Transactions on Information Systems, and IEEE Intelligent Systems, and was the program co-chair of HLT-NAACL '03 and SIGIR '99. She has received an NSF CAREER award, an IBM Faculty Award, a Google Research Award, an Okawa Foundation Fellowship, and two student-initiated Excellence in Teaching awards.

This was published as a guest article in Stephen Few's monthly Visual Business Intelligence Newsletter. A complete library of Stephen Few's articles, as well as other guest articles, is available at www.perceptualedge.com.