Usability News

April 2009, *Vol. 11 Issue 1*

| **Volume 11 Issue 1** | **Past Issues** | **A-Z List** |

Using Tag Clouds to Visualize Text Data Patterns

Veronica Hinkle

Summary. This article provides an overview of the use of tag clouds to represent textual content. A tag cloud is a visual representation of textual content that uses size and color to indicate word frequency. Several examples of tag clouds are shown and the advantages and disadvantages of using them to summarize textual data for website design are discussed.

INTRODUCTION

Tag clouds are visual representations that indicate frequency of word usage within textual content, which include web sites, articles, databases, or speeches. The frequency of each word or "tag" is represented in the tag cloud by increasing the font size and color saturation of that word.

Tag clouds are traditionally used to represent the frequency of tags, which can be words within the content or descriptions of the content, created by users from social software (Rivadeneira, Gruen, Muller, & Millen, 2007) and social bookmarking websites (Al-Khalifa, Davis, & Gilbert, 2007; Hassan-Montero & Herrero-Solana, 2006). These user-created website descriptions are also called folksonomies, which are personal categorizations of a website's content and purpose. A folksonomy is a personal taxonomy of descriptions for a content source from the perspective of the user (Al-Khalifa et al., 2007; Guy & Tonkin, 2006).

Recently, tag clouds have been shown on web sites as a means to summarize content. The following are examples, which have been acknowledged both by cyberspace bloggers and researchers, who have implemented tag clouds to aid users in navigating and browsing information on their domains.

- **Flickr**. This is a photo management website where users can host, manage, and share pictures with the public. It was one of the first web sites to use tag clouds to help users navigate photo collections hosted on the site. The tag cloud in Figure 1 is used to help users explore different photo themes. Each word is a hyperlink.
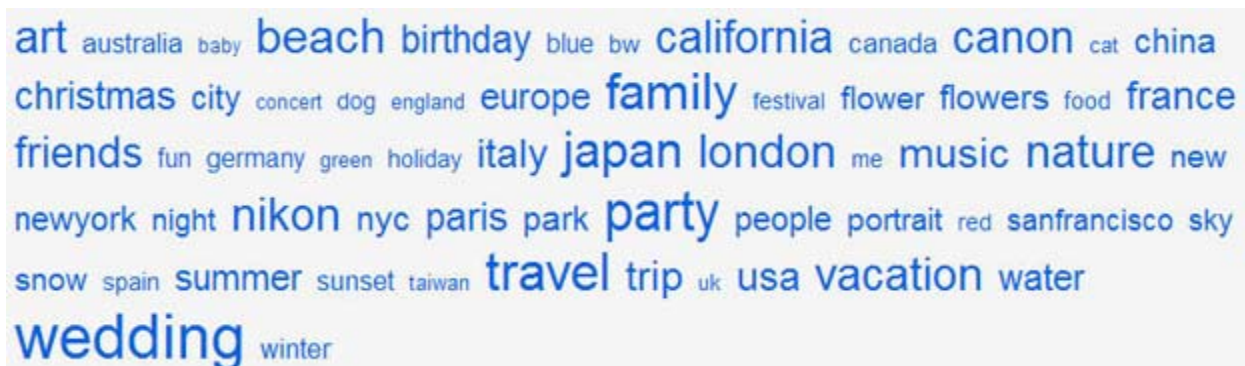
Figure 1. Tag Cloud from Flickr

- **Delicious**. This is a social bookmarking website where users can bookmark websites online and browse the Internet based on previous tags representing popular bookmarks. Unlike other tag clouds, Delicious' tag clouds follow a vertical list layout and provides a cloud option under the tag options. Figures 2 and 3 show the tags users see when searching for information. In this case, the tag "design" was selected. Figure 3 shows the tags for the "design" tag clicked on Figure 2.
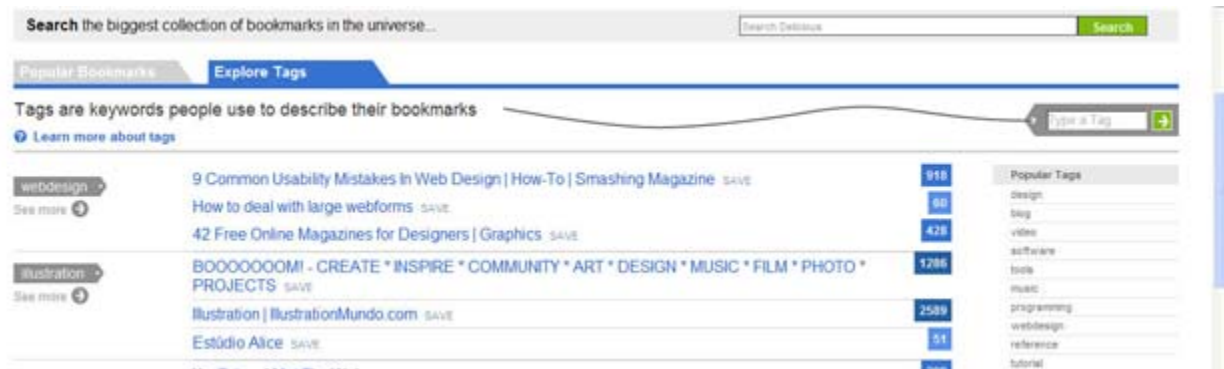
Figure 2. Tag Cloud from Delicious
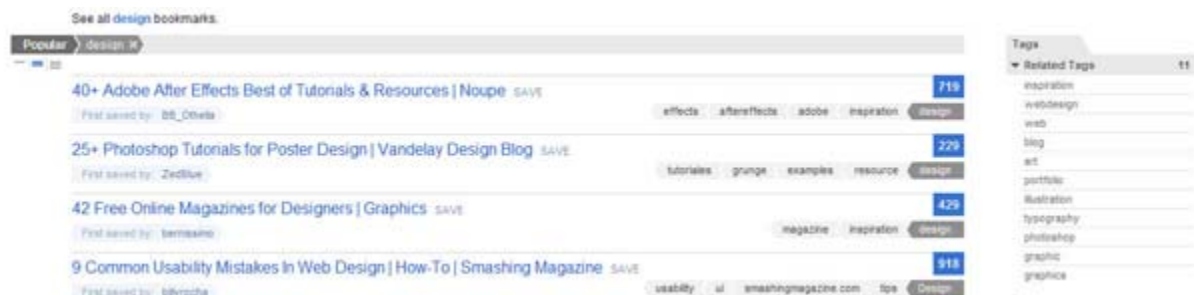
Figure 3. Tags for Previously Selected Tag from Delicious

- **Pollster**. This is a website that publishes poll results and political commentaries. Figure 4 shows a tag cloud from Pollster as a way to analyze and compare speeches from the democratic debate on April 26, 2007.
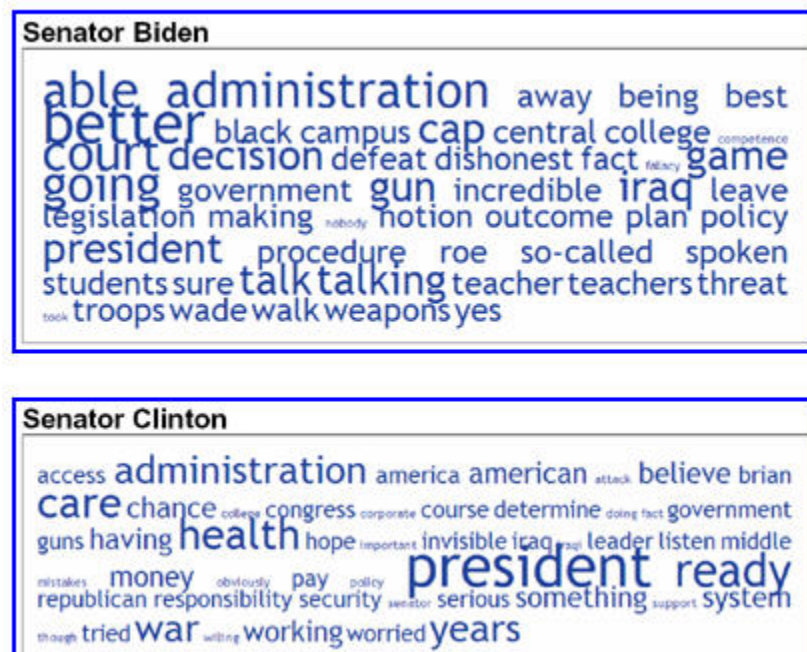
Figure 4. Tag Clouds from the [Democratic Debate from Pollster](#)

Information Architecture researchers have developed more sophisticated algorithms to create Tag Maps, a way to visualize large collections of geo-referenced photos (Jaff, Naaman, Tassa, & Davis, 2006). The result is text tags, which are placed on a map of a selected geographic region. The tag is displayed at the location where it takes place and the size of the tag's text represents its importance.

Another impressive new implementation of tag clouds is the idea of using multiple synchronized tag clouds to support user exploration of semi-structured clinical trial data by using the CTSearch application, which is a search tool for [ClinicalTrials.gov](#) (Hernandez, Falconer, Storey, Carini, & Sim; 2008). The idea is to supplement traditional ways to search for data by providing tag clouds as visualizations of the semi-structured data in a form that would allow the user to scan the important terms highlighted by the tag clouds and relate that information to the query results.

There are four types of tasks that tag clouds can support: search, browsing, recognition and matching, and impression formation or gisting (Rivadeneira et al., 2007). This article focuses on impression formation or gisting text information from tag clouds for analysis purposes.

Analyzing Text responses use tag clouds

As the Pollster's tag clouds of political speeches show, tag clouds can help researchers visualize response patterns from text responses. It is convenient to use tag cloud generators to create these visualizations to aid analyzing text responses and discourse. Although it is not advocated as a substitute for traditional techniques used for analyzing text responses, using tag clouds to glance at the overall patterns of text responses can be extremely beneficial to see the semantic patterns or gist of a particular set of answers.

The following are some examples of innovative ways to use tag clouds.

- **Responses from an open-ended question in a survey**. Figure 5 shows a tag cloud of responses to an open-ended survey question regarding concerns about tiger exhibits in a zoo. It can be seen from this visualization that the respondents were concerned about the safety of people and livestock near the exhibit.

Figure 5. Responses from an open-ended question from a survey.

- **Looking at group names generated from a card sorting exercise**. Although frequency counts and dendograms are traditionally used to interpret card sorting results, analyzing suggested group names can be done by simply "eyeballing" the data or doing simple frequency counts of the group names though tag cloud analysis. At SURL, we have used tag clouds to visualize user-suggested group names for a card sort of 100+ items. This helped us to determine which group names, or terms, were the most commonly suggested by participants. Figure 6 shows suggested group names for a list of cell phone terms that were part of the card sorting exercise.

Figure 6. Suggested group names for a card sorting of cell phone functions.

- **Developing a list of terms for a card sorting exercise**. We also have used tag clouds to analyze suggested names for descriptions of specific cell phone functions. The tag clouds facilitated our group discussions regarding what terms to include as part of a subsequent card sort exercise. Figure 7 shows the generated tag cloud for the definition: "What term would you use to describe 'hold my services while I am on vacation.'"

Figure 7. Tag cloud of user-suggested terms for "hold my services while I am on vacation"

- **Showing themes from a text collection**. Cooke (2008) used a tag cloud to visualize the most frequently used terms found in article titles from the Human Factors and Ergonomics Society's (HFES) journal over the past 50 years. Figure 8* shows how the tag cloud helps highlight the themes of the journal's publications.

Figure 8. Tag cloud of frequently used terms in articles' titles for the HFES journal.

Choosing a Tag Cloud Application

Several tools can be used to generate tag clouds. The usability of these tools in generating effective and efficient tag clouds can be compromised by different factors, such as how the tags are organized and how the most frequent tags are represented. The following list presents research findings from studies focusing on factors that affect the overall usability of tag clouds as visualization tools.

- Font size has a strong effect on users trying to locate a tag. The layout of tags within the cloud affects impression formation or gisting. Tags that are ordered by frequency have the best identification (Rivadeneira et al., 2007).
- Alphabetization of the tags aid users in finding information faster. Larger font sizes seems to decrease time to find a tag and makes the process easier. Users tend to scan the tag clouds rather than read them when trying to locate information (Halvey & Keane, 2007).
- Certain visual features of tag clouds, such as font size, font weight, saturation, and color, influence how fast we can find tags in the cloud. In the case of color, it is important to take into account how color can increase the saliency of a tag inadvertently. Features that do not have strong visual influence effects include number of pixels, width, and area (Bateman, Gutwin, & Nacenta; 2008).

Given these considerations, it is important to select a tag cloud application that emphasizes ease of tag retrieval and makes the process of generating a tag cloud simple and efficient.

Tag Cloud Applications

The following is a list of free applications for creating tag clouds and a sample tag cloud from each. A list of suggested group names generated from a card sort study of cell phone information architecture was used to create the tag clouds in each of the applications listed. It can be seen that the visual representation varies greatly across applications. Applications that were for web purposes only were not included.

- Tagcrowd

- ManyEyes

- [Wordle](#)

- [MakeCloud](#)

- [Tag Cloud Builder](#)

CONCLUSION

The use of tag clouds can greatly aid the analysis of text responses in many situations. It is important to choose the tag cloud application carefully and understand how the visual effects of the tag clouds can bias the interpretation of the visualization. Equally important is to take into account how the tag cloud is generated, what words are being left out, and how many tags are being created per cloud. When choosing a tag cloud application, it is important to consider how much control you have over the features to generate the tags, such as how words are counted, how words are omitted, how different font sizes are estimated, and how similar words are scored.

REFERENCES

Al-Khalifa, H. S., Davis, H. C. and Gilbert, L. (2007). Creating Structure from Disorder: Using Folksonomies to Create Semantic Metadata. 3rd International Conference on Web Information Systems and Technologies (WEBIST), 3 - 6 March, 2007, Barcelona, Spain.

Bateman, S.; Gutwin, C. & Nacenta, M. (2008). Proceedings of the nineteenth ACM conference on Hypertext and hypermedia, 193-202.

Cooke, N. (2008). Celebrating 50 Years of Human Factors. Human Factors and Ergonomics Society Bulletin, 51(5), 1-3.

Guy, M. & Tonkin, E. (2006). Folksonomies: Tidying up Tags? D-Lib Magazine.

Halvey, M. & Keane, M. (2007). An Assessment of Tag Presentation Techniques. Proceedings of the 16th international conference on World Wide Web, pages 1313–1314.

Hassan-Montero, Y. & Herrero-Solana, V. (2006). Improving Tag-Clouds as Visual Information Retrieval Interfaces. http://www.nosolousabilidad.com/hassan/improving_tagclouds.pdf /International Conference on Multidisciplinary Information Sciences and Technologies. Accessed Wednesday, February 18, 2009.

Hernandez, M.; Falconer, S.; Storey, M.; Carini, S.; Sim, I. (2008). Synchronized Tag Clouds for Exploring Semi-structured Clinical Trial Data. IBM Centre for Advanced Studies Conference. Proceedings of the conference of the center for advanced studies on collaborative research: meeting of minds, 1-15.

Jaffe, A.; Naaman, T.; Tassa, T.; Davis, M. (2006). Generating Summaries and Visualization for Large Collections of Geo-referenced Photographs. Proceedings of the 8th International Multimedia Conference, 89-98.

Rivadeneira, A.; Gruen, D.; Muller, J.; Millen, D. (2007). Getting our Heads in the Clouds: Toward Evaluation Studies of Tag Clouds. CHI 2007 Proceedings.

http://www.joelamantia.com/blog/archives/ideas/tag_clouds_evolve_understanding_tag_clouds_1.html. Accessed February 18, 2009.

http://www.smashingmagazine.com/2007/11/07/tag-clouds-gallery-examples-and-good-practices. Accessed February 18, 2009.

http://www.tagcrowd.com. Accessed February 18, 2009.

http://manyeyes.alphaworks.ibm.com/manyeyes/visualizations. Accessed February 18, 2009.

http://www.wordle.net. Accessed February 18, 2009.

http://www.flickr.com/explore. Accessed February 18, 2009.

SUBSCRIBE to *Usability News*!