Above the Tag Clouds: Visualizing the Temporality of Tags

Abstract

Tag clouds are becoming increasingly popular with websites that utilize social tagging to categorize ever expanding collections of digital information. Tagging has been found to be more adaptable than traditional classification, as well as more prone to serendipitous information discovery. The flexibility of tagging systems allows users to rapidly adopt new terms and engage in extremely dynamic tagging practices, yet tag clouds are not able to represent agile shifts in tagging patterns. Over time, semantic and linguistic changes can modify the meaning and form of tags, and changes in tagging behavior can create disconnects between related tags. By conceiving tagging as a triad: object, user, tag, we completely miss the critical notion of time. Time leads to changes in semantics, vocabulary, behavior, and syntax. In order to address the problem of aging tags and aging folksonomies, we really need to include time as a critical facet of tagging: object, user, tag, time. The adaptive behavior of tags requires that there is a constant influx of new descriptive data about an object, but time-related changes have to overcome the weight of the pre-existing tags. In this poster we propose a new tag-cloud visualization technique that attempts to address these issues by including a dynamic factor: the changing weight of tags over time.

Background

The number of websites that support tagging has rapidly increased since 2003, with sites such as Del.icio.us, Flickr.com, Technorati.com, and Librarything.com, gaining increasing popularity. Each of the sites is specialized, allowing users to store, organize, and later retrieve specific digital resources. For example, Del.icio.us was designed for the sole purpose of tagging URLs, Flickr.com allows tagging of photographs, Technorati.com tracks the tagging of blogs, and Librarything.com of personal book collections. These tags produce a flat namespace, meaning that there is no hierarchy among terms, and no specified parent-child relationship between terms. Although the flat namespace frees users from a rigid structure, it also creates complexity around resource retrieval. Tools such as tag clouds - which visually display tags by frequency - attempt to aid retrieval through a compact representation that "draws the eye towards the largest, and presumably most important items," (Hearst and Rosner 2008).

According to Rivadeneira et. al. (2007), there are two types of features that can be used to construct a tag cloud: text and placement:

- Text Features: font weight, size, color,intensity, text width
- Placement: word sorting, clustering and spatial layout

When combined, the features can aid in a range of tag cloud tasks such as: searching for a specific ter, browsing the tag cloud, impression formation, and tag recognition and matching (Rivadeneira et. al. 2007; Bateman 2007). However, the majority of tag clouds only utilize varying font sizes and weights to visually differentiate between frequency of term use. Terms that are used more frequently are presented in a large font size, while less frequent terms are minimized. There are several problems with this type of implementation:

- Difficulty comparing tags with a similar size
- Difficulty seeing smaller tags when they are located next to larger tags
- The length of the word might be conflated with its size, making it seem more important
- The most frequently used tags are the most general tags
- Tag clouds are normally presented in alphabetical order so there is a lack of information about relationships between tags
- There is no visual flow through the display which causes the eye to dart around (Bateman 2007; Hearst and Rosner 2007; Haynes et.al. 2007; Michlmayr and Cayzer 2007).

In order to solve some of the problems with tag clouds listed above, researchers have been experimenting with combining placement along with text. A number of examples are shown below.

Problems Related to Tag Aging

As previously discussed, tagging systems allow users to rapidly adopt new terms and engage in extremely dynamic tagging practices, yet tag clouds are not able to represent agile shifts in tagging patterns. Over time, semantic and linguistic changes can modify the meaning and form of tags, and changes in tagging behavior can create disconnects between related tags.

Semantic changes are changes in the meaning of a word over time. There are a number of ways in which this happens; words become more (or less) specific, words change from positive to negative, and words change in meaning based on their similarity to another concept.

In addition to semantic changes, there are a number of other linguistic shifts that cause problems within tagging systems:

- Spelling Changes
- Syntactic Changes – Changes in punctuation and grammar
- Neologism – New words are coined (ex. AJAX)

Tagging behavior can change over time, affecting a users ability to re-find information tagged using now obsolete patterns and behaviors.

- Users gain experience, which leads to changes in tagging behavior (Vanderwal 2007)
- Users modify their tagging behavior based on external feedback and pressures (Sen et. al. 2006)

Finally, studies of tag vocabularies have noted that tags generally follow power-law growth curves, with the number of new tags diminishing over time (Cattuto et. al. 2007, Golder and Huberman 2006). This growth pattern leads to a steady state where the proportionate usage of tags stabilizes over time. In effect, tag clouds will stabilize over time, and will resist changes in tagging behavior, sematics, and form.

Agile Tag Clouds

A number of visualization and analytical methods have been proposed that address the problems of tag aging. For example: Hassan-Montero and Herrero-Solana (2006) used clustering techniques to identify tags with stronger discriminatory value, Kuo et. al. (2007) used stemming to combine tag variations and color variations to identify recently used tags.

In addition, studies of tagging systems have shown "bursts" of activity (Golder & Huberman 2006) even after tag vocabularies have stabilized. These bursts may be rooted in any of the disruptive effects we previously discussed.

In order to make tag clouds more adaptable to these bursts — and building on previous efforts to adapt tag clouds to the effects of aging — we propose a new tag cloud visualization method that incorporates change in the descriptive power of a tag over time. In this method we include three factors in determining the weight of a given tag within the cloud:

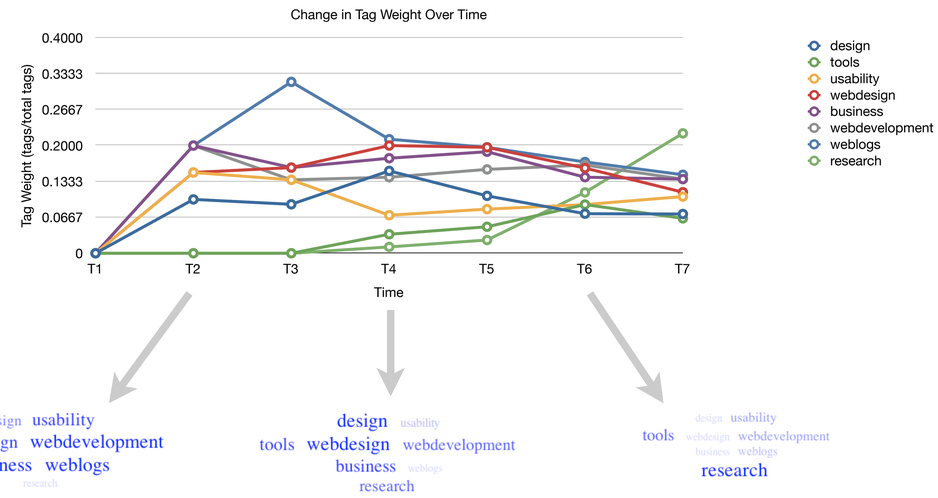
- Tag Frequency — The relative frequency in a given tag
- Tag Age — The amount of time elapsed since a tag was last used
- Tag Frequency Change — The change in relative frequency over time

We separate these into two groups: Time (including tag age and tag usage rate) and Weight (including tag frequency). The first group, Time, is visualized by changes in the color intensity of a tag: tags that are gaining in usage are more intense, while tags that are no longer being used are given a less intense color. The second group is visualized by changes in the size of the tag — the higher the frequency of the tag, the bigger it is within the tag cloud.

We are utilizing co-occurence clustering methods similar to Hassan-Montero and Herrero-Solana (2006) to group semantically-related tags, and sorting tags within those groups to aid in visual scanning.

Agile Tag Cloud Example

This example shows that how the agile tag cloud method allows for tag clouds to quickly adapt to changes in tagging behavior.



T2: Several tags are growing in usage.

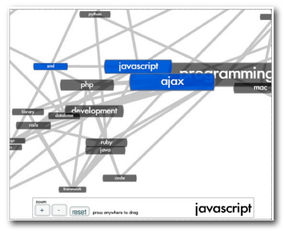
T4: We see that the usability tag has lost ground and design is being used more frequently

T6: The tag cloud is stabilizing, but the research tag is quickly growing in usage.

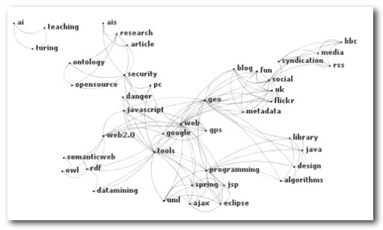
A "classic" tag cloud from Flickr.com. Tags are weighted by frequency.

Hassan-Montero and Herrero-Solana (2006), utilized a clustering algorithm based on co-occurrence analysis to come up with the following visualization:

Bielenberg and Zacher (2006) presented their tag cloud in circular form, where font size and distance to the center represent the importance of the tag.



Shaw (2005) and Michlmayer and Cayzer (2007) proposed to map a tag cloud to a network graph, where tags are represented as nodes in the graph and similarity among tags as edges.



References

Bateman, S. (2007). *Collaborative Tagging: Folksonomy, Metadata, Vizualization, E-Learning*. Unpublished master's thesis, University of Saskatchewan.

Cattuto, C., Loreto, V., and Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. PNAS, 104(5):1461-1464.

Golder, S., and Huberman, B. (2006). Usage patterns of collaborative tagging systems. Journal of Information Science, 32(2), 198-208.

Halvey, M., and Keane, Mark (2007). An Assessment of Tag Presentation Techniques. *WWW07 Poster Paper*, May 8-12, 2007, Banff, Alberta, Canada.

Hassan-Montero, Y., and Herrero-Solana, V. (2006). Improving Tag-Clouds as Visual Information Retrieval Interfaces. *International Conferene on Multidisciplinary Information Sciences and Technologies*, October 25-28, 2006, Merida, Spain.

Hearst, M., and Rosner, D. (2008). Tag Clouds: Data Analysis Tool or Social Signaller? *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, 7-10 Jan. 2008.

Kuo, B. Y-L., Hentrich, T., Good, B. M., Wilkinson, M. D. (2007). Tag Clouds for Summarizing Web Search Results. *WWW07*, May 8-12, 2007, Banff, Alberta, Canada.

Michlmayr, E., and Cayzer, S. (2007). Learning User Profiles from Tagging Data and Leveraging them for Personal(ized) Information Access. *WWW07*, May 8-12, 2007, Banff, Alberta, Canada.

Rivadneira, A.W., Gruen, D., Muller, M., and Millen, D (2007). Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds. *CHI 2007 Proceedings*, April 28-May 3, 2007, San Jose, CA.

Sen, S., Lam, S., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., Riedl, J. (2006). tagging, communities, vocabulary, evolution. *CSCW'06*, Nov. 4-8, 2006, Banff, Alberta, Canada.

Future Work

We are currently implementing a prototype tag cloud tool that uses the method described in this poster for visualizing tag sets. This prototype will be used for future user testing. We are also interested in exploring variations on these methods, including the use of tag weight in place of tag frequency when looking at community tag clouds.

Jackson Fox (jacksonfox@gmail.com), Lulu.com
Alla Zollers (azollers@gmail.com), UCLA