Patterns in the Clouds - The Effects of Clustered Presentation on Tag Cloud Interaction

Johann Schrammel[1(✉)] and Manfred Tscheligi[2]

[1] CURE – Center for Usability Research and Engineering, Vienna, Austria
schrammel@cure.at
[2] ICT&S Center, University of Salzburg, Salzburg, Austria

Abstract. Tag clouds have become a frequently used interaction technique in web-based systems. Recently, different clustered presentation approaches have been suggested to improve usability and utility of tag clouds. In this paper we describe a modified layout strategy for clustered tag clouds and report the findings of an empirical evaluation of automatically clustered tag clouds with 22 participants for both specific and general search tasks. The evaluation showed that automatically clustered presentation performs as well as alphabetic layouts in specific search tasks and that clustered presentation is an improvement over random layout for general search tasks. Clustered tag cloud presentation also was preferred by a majority of users for general search tasks. High quality of the clustering was mentioned as key variable for usefulness of the approach in the qualitative interviews with the users.

Introduction

Since the first introduction of tag clouds different means to further enhance their usefulness have been proposed. Suggestions for modifications were directed towards the utilization of additional display properties for encoding more data dimensions [15], the optimization of the layout algorithms [17], adaptation of sorting strategies (e.g. alphabetically versus importance-based) [13], or the combination of tags with graphical elements [9]. A specifically popular direction of research has been along the lines of clustering and displaying tags along their semantic meaning, and different approaches have been suggested [2, 3, 5].

Only few empirical evaluations exist assessing the expected advantages, and were they are available they did find no or only minor advantages [7, 11]. We think that these rather discouraging results are partly due to shortcomings in the used clustering methods and presentation approaches of semantically clustered tag clouds. Typically the used methods are not optimized for the most relevant tasks and context situations.

Another critical element regarding the usefulness of clustering approaches for use in tag cloud display is the quality of the automatically calculated clusters. Evaluations of human-made clusters based on hand-picked data have shown very promising results for usage of clustered approaches [7]. Results for methods that use automated clustering however have been much less convincing [11]. The quality of the clustering

algorithm and whether the resulting clusters are understandable for humans seem to be of major importance with regard to the usefulness of clustered presentation approaches.

Also, the type of task a user is working on has been shown to be a main influence on whether an interface solution is perceived well by users or not [13]. Therefore in our work we address both specific and general search tasks.

In our work we want to answer the question whether similar results as with hand-made clusters could be achieved with realistic data and state-of-the-art clustering algorithms. We developed a rectangular clustered layout approach and evaluated it in the context of specific and general search tasks. In the next sections we present related work, the study design and the evaluation results.

Related Work

Visual Features of Tag Clouds

The importance of visual features of tags within tag clouds for attention has been researched recently, and results from different authors [1, 10] show that font size, font weight and intensity prove to be the most important variables. Regarding the importance of tag position reported empirical findings are not as concise. Whereas [1] found no influence of tag position other researchers [7, 10, 11] report that tags in the upper-left quadrant receive more attention than tags in the lower-right quadrant. Tag clouds and information seeking tasks.

Sinclair et al. [13] compared the usefulness of tag clouds against search interfaces for general and specific information seeking tasks and concluded that tag clouds are especially useful for non-specific information discovery as they can provide a helpful visual summary of the available contents and its relevance. Similarly, comparing the visualization of search results using tag clouds in contrast to hierarchical textual descriptions Kuo et al. [9] found that users were able to answer overall questions better when using tag clouds. Regarding specific search tasks however both studies showed disadvantages for tag clouds. Using eye tracking data to analyze the effect of introducing search results overview in the form of a tag cloud Gwizdka and Cole [18] found that a results overview in form of a tag cloud helps a user to become faster and more efficient.

Layout of Tag Clouds

Halvey and Keane [4] investigated the effects of different tag cloud and list arrangements comparing the performance for searching specific items. The setup included random and alphabetically ordered lists and tag clouds. Clustered presentation was not part of their setup. They found that respondents were able to more easily and quickly find tags in alphabetical orders (both in lists and clouds).

Rivadeneira et al. [10] compared the recognition of single tags in alphabetical, sequential–frequency (most important tag at the left-upper side), spatially packed (arranged with Feinberg's algorithm, for more information see www.wordle.net) and

list-frequency layouts (most important tag at the beginning of a vertical list of tags). Results did not show any significant disparity in recognition of tags. However, respondents could better recognize the overall categories presented when confronted with the vertical list of tags ordered by frequency.

Hearst and Rosner [6] discuss the organization of tag clouds. One important disadvantage of tag cloud layouts they mention is that items with similar meaning may lie far apart, and so meaningful associates may be missed.

Semantic Tag Clouds

Hassan-Montero and Herrero-Solana [5] proposed an algorithm using tag similarity to group and arrange tag clouds. They calculate tag similarity by means of relative co-occurrence between tags. Likewise, Fujimura et al. [3] use the cosine similarity of tag feature vectors (terms and their weight generated from a set of tagged documents) to measure tag similarity. Based on this similarity they calculate a tag layout, where distance between tags represents semantic relatedness. Another very similar approach is proposed by [2].

Semantic approaches have been evaluated recently by different researchers. Schrammel et al. [11, 12] evaluated a semantic layout approach that places related tags together but does not explicitly calculate and present groups of tags. They report that semantic layouts can provide minor advantages, and that it was difficult for users to identify and understand the layout strategy.

Lohmann et al. [7] studied a clustered layout were groups of similar tags were placed together and indicated by border lines and background shading. They report advantages of the clustered layout for general search tasks. However, as they used a manually constructed tag corpus and provide no details on how the clustering was calculated the question remains whether these results can be replicated with realistic data and unsupervised clustering algorithms.

Research Questions

In detail we wanted to answer the questions how automatically clustered tag layouts affects search time, the perception of tag clouds as well as the subjective satisfaction of the users after interacting with the tag clouds both when searching for a specific tag and when performing searches for tags that belong to a specific topic. We compare three layout strategies: alphabetic (the currently most used approach), random (to be able to see if clustered presentation provides any improvement over no structure at all) and automatically clustered.

Study Materials and Participants

Tag Corpora. As a basis for our work we decided to use data from del.icio.us, as this site allows everybody to tag and that the site employs a blind tagging process i.e. the

users cannot see which tags where used by other users during the tagging process. In detail our work is based on a large data sample that was downloaded from 'del.icio.us' by Yusef Hassan-Montero, who thankfully provided us with the data. The data originally was collected for research described in detail in [5]. Data was crawled by means of an automatic crawler during October 2005 and contains 218,063 URLs tagged with 242,349 tags by 111,234 users.

Clustering. To calculate tag similarity we used a well proven method known as Jaccard coefficient. Similarity between tags is measured by the intersection divided by the union of the sample set. Based on this similarity measure clusters of tags where calculated using the bisecting k-means approach. For a discussion of different clustering approaches and their pros and cons see Steinbach et al. [14]. The clusters were calculated using the CLUTO-Toolkit provided and described by [8]. Basically the N-dimensional similarity matrix of tags was used as an input for the clustering algorithm. The target number of clusters to calculate was specified as 20. This number was chosen to form clusters of about five tags, which informal pre-test showed to be a good size for clusters.

Tag Selection for Test Content. Six different tag content sets were needed to guarantee that participants worked with a new content set in every condition. To construct the different tag content sets the 600 most useful tags according to the improved selection mechanism described by [5] were chosen from the delicious data set. Tags where then divided into three groups according to their frequency of use. This later one is used to decide on the size of the item in the tag clouds. The three different groups were not of equal size, as this would result in an unaesthetic and inefficient use of tag clouds.

Tag Cloud Composition. Next, each of these three tag collections was divided into groups of six items to form the basis for the different needed tag clouds. Tags where assigned to groups starting from the tags with the highest value for usefulness continuing to the lower values (again based on [5]). Then the tags of these groups were assigned randomly to the six test content sets. With this procedure we could ensure both that (a) all tag clouds have the same number of big, medium and small tags and that (b) the items of the different tag clouds are of similar quality and usefulness.

Tag Cloud Design. In contrast to [5] who place each cluster in a new line or [3] who translate semantic distance into screen distances we decided to keep the typically used rectangular layout of tag clouds. The reason for this approach is its efficiency with regard to screen real estate, and the advantages regarding readability and scanability. Furthermore this design layout eases implementation.

To mark the different clusters color-coding was used: each tag within a cluster was underlined with the same color. To avoid disconnecting clusters spatially in case of line-breaks tags were placed in the tag cloud in a zig-zag-manner (i.e. one line was filled in the normal reading direction left-to-right with tags, and the next from right to left). Clusters were also separated from each other by additional blank space to enhance immediate perceptibility of clusters. The placement approach also reflects the similarity between clusters and not only between tags. Clusters that are placed near

(a)   Alphabetic

(b)   Clustered

Fig. 1. Example content displayed in alphabetic and clustered layout condition

each other are more similar than clusters with great distance between them. Figure 1(b) shows an example from the set of constructed clustered tag clouds.

Participants. 22 user (17 male, 5 female) participated in the evaluation. Average age of participants was 31.9 years (Min: 25, Max: 53). All of them had normal or corrected to normal vision. All participants had a technical background (because of the used tag corpora from delicious which contains many technical terms) and were intense users of web technologies.

Experiment One: Finding Specific Tags

The first experiment was designed to test how clustered tag layout influences search time and subjective evaluation of task difficulty when searching for a specific tag within a tag cloud. The task for the test participants was to find a predefined tag within a tag cloud as fast and accurately as possible.

The tag to be found was shown on the screen, on clicking 'Next' a tag cloud containing the target word appeared on the screen. The target word was also shown below the tag cloud. After locating the target tag participants had to click on it to proceed to the next task. Search time and clicked tag was logged.

For each layout condition twelve search tasks for different targets within the same tag cloud where performed. Target tags where evenly distributed across the three font sizes. We controlled for evenly distributed target position across the four quadrants of

the clouds used in each condition, as prior research showed that tag position can have relevant influence [7, 10]. Presentation order of layout during the test procedure was systematically varied to counterbalance possible order effects.

Effects of Tag Cloud Layout on Search Time

Repeated Measures Analysis of Variance showed a significant influence of the layout condition on search time ($F_{2,42} = 61.48$, $p < 0.000$). Post-hoc analysis using paired-samples t-tests with Bonferroni-corrected alpha levels showed that the alphabetic layout is significantly faster than random ($t_{21} = -12.4$, $p < 0.000$) or clustered ($t_{21} = -10.3$, $p < 0.000$) layout. There is no significant difference between random and clustered layout ($t_{21} = 0.0$, $p = 0.33$). Average search times are shown in Table 1 below.

Table 1. Mean search times in seconds for the three layout strategies for specific searches (Experiment One) and general searches (Experiment Two)

	Alphabetic	Clustered	Random
Specific search (Exp. One)	3.7 s	13.3 s	14.6 s
General search (Exp. Two)	18.6 s	17.1 s	22.0 s

Experiment Two: Finding Tags Related to a Specific Topic

In the second part of the study the task of the participants was to find a specific tag that belongs to a pre-defined category. The categories were selected manually by the researchers. Special care was given that only unambiguous categories were used. Table 2 below provides examples of tasks for the selected categories for the tag cloud shown in Fig. 1. All categories were verified by informal testing with colleagues of the authors with regard to their understandability and unambiguousness.

For every tag cloud three categorical search tasks were defined that contained multiple (two or three) relevant tags that were grouped together by the clustering algorithm into one cluster. Similarly, two categorical search tasks were defined that also contained two tags, but where the clustering algorithm had placed these tags into different clusters. Furthermore ten tasks were specified, where only one correct target tag existed. Again special care was given that these target tags where evenly distributed across all quadrants in the alphabetic, random and in the grouped tag cloud layout. Table 2 below shows example tasks for all three task categories.

Table 2. Example tasks for general search in Experiment Two

Category	Task	Solution
Multiple tags in same cluster	City in the USA	Seattle, Sanfrancisco
Multiple tags in different clusters	IT-Enterprise	Sun, apple, ebay
Only one target	Name of a continent	Europe

Effects of Tag Cloud Layout on Search Time

Repeated Measures Analysis of Variance showed a significant influence of the layout condition on search time ($F_{2,42} = 3.37$, $p = 0.044$). Post-hoc analysis using paired-samples t-tests with Bonferroni-corrected alpha levels showed that the clustered layout is significantly faster than the random layout ($t_{21} = 2.6$, $p = 0.017$). Even though mean search time for clustered layout is 1.5 seconds faster than for alphabetic this difference is not statistically significant ($t_{21} = 0.96$, $p = 0.349$). Based on information from the qualitative interviews we think this is due to the very high variation in the data which is caused by cases were test persons did overlook a tag and had to scan the tag cloud for very long time.

User Preferences

After conduction of the experiment users were asked to state their preference for a layout strategy both when searching for a specific tag and when trying to achieve an overview on a web page. All except one participant preferred alphabetic layout for the specific search. For gaining orientation and overview a majority of users preferred the clustered layout (15) over alphabetic (4) or random (3) layout strategies.

Qualitative Comments of Users

After each experiment users were briefly interviewed regarding their subjective impression regarding the clustered presentation approach. The general impression can be summarized as positive. Most participants really liked the approach for orientation tasks and general searches. Almost everyone also mentioned having been irritated and confused by some arbitrary looking clusters or 'wrong' placements of tags. Another negative aspect mentioned was the additional cognitive cost for understanding the meaning of a cluster. Few participants were irritated by the colors used to mark the clusters.

Discussion and Conclusion

Clustered tag cloud layouts seem to have the potential to improve search performance and satisfaction for general search tasks. However, our results (especially from the qualitative interviews) also show that state-of-the-art clustering mechanisms still produce artifacts that are difficult to understand by the users, and that counteract the possible usefulness of the approach. Application of clustered approaches therefore is only recommended in case sufficient quality of the clustering can be ensured. Results for specific searches show - as expected - that clustered presentation is only suited for application contexts were the main goal of the users is to gain an overview, and were searching for specific contents is secondary.

We could show that clustering tags in tag clouds is feasible in realistic settings i.e. using real data and applying state-of-the art clustering algorithms, and produces satisfactory results that are welcomed by users for general searches.

In future we plan to work on tackling the problems arising from suboptimal clusters. We want to explore the effects of only marking clusters with high internal homogeneity, and to use machine learning based categorization approaches to be able to also label found clusters.

References

1. Bateman, S., Gutwin, C., Nacenta, M.: Seeing things in the clouds: the effect of visual features on tag cloud selections. In: Proceedings of Hypertext and Hypermedia 2008, pp. 193–202. ACM Press, New York (2008)
2. Berlocher, I., Lee, K., Kim, K.: TopicRank: bringing insight to users. In: Proceedings of SIGIR 2008, pp. 703–704. ACM Press, New York (2008)
3. Fujimura, K., Fujimura, S., Matsubayashi, T., Yamada, T., Okuda, H.: Topigraphy: visualization for large-scale tag clouds. In: Proceedings of WWW 2008, pp. 1087–1088. ACM Press, New York (2008)
4. Halvey, M.J., Keane, M.T.: An assessment of tag presentation techniques. In: Proceedings of WWW 2007, pp. 1313–1314. ACM Press, New York (2007)
5. Hassan-Montero, Y., Herrero-Solana, V.: Improving tagclouds as visual information retrieval interfaces. In: Proceedings of InfoSciT (2006)
6. Hearst, M.A., Rosner, D.: Tag clouds: data analysis tool or social signaller? In: Proceedings of HICSS (2008)
7. Lohmann, S., Ziegler, J., Tetzlaff, L.: Comparison of tag cloud layouts: task-related performance and visual exploration. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) Human-Computer Interaction-INTERACT 2009. LNCS, vol. 5726, pp. 392–404. Springer, Heidelberg (2009)
8. Karypis, G.: CLUTO - a clustering toolkit. Technical Report #02–017, November 2003
9. Kuo, B.Y., Hentrich, T., Good, B.M., Wilkinson, M.D.: Tag clouds for summarizing web search results. In: Proceedings of World Wide Web 2007, pp. 1203–1204. ACM Press, New York (2007)
10. Rivadeneira, A.W., Gruen, D.M., Muller, M.J., Millen, D.R.: Getting our head in the clouds: toward evaluation studies of tagclouds. In: Proceedings of CHI 2007, pp. 995–998. ACM Press, San Jose (2007)
11. Schrammel, J., Leitner, M., Tscheligi, M.: Semantically structured tag clouds: an empirical evaluation of clustered presentation approaches. In: Proceedings of CHI '09, pp. 2037–2040. ACM, New York (2009)
12. Schrammel, J., Deutsch, S., Tscheligi, M.: The visual perception of tag clouds - results from an eye tracking study. In: Proceedings of Human-Computer Interaction - INTERACT 2009, 12th IFIP TC 13 International Conference, Uppsala (2009)
13. Sinclair, J., Cardew-Hall, M.: The folksonomy tag cloud: when is it useful? J. Inf. Sci. **34**(1), 15–29 (2008)
14. Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. In: Grobelnik, M., Mladenic, D., Milic-Frayling, N. (eds.) KDD-2000 Workshop on Text Mining. Boston (2000)

15. Waldner, M., Schrammel, J., Klein, M., Kristjansdottir, K., Unger, D., Tscheligi, M.: FacetClouds: exploring tag clouds for multi-dimensional data. In: Proceedings of the Graphics Interface Conference, pp. 17–24 (2013)
16. Lee, B., Riche, N.H., Karlson, A.K., Carpendale, S.: Sparkclouds: visualizing trends in tag clouds. IEEE Trans. Vis. Comput. Graph. **16**(6), 1182–1189 (2010)
17. Kaser, O., Lemire, D.: Tag-cloud drawing: algorithms for cloud visualization. In: Proceedings of Tagging and Metadata for Social Information Organization (WWW 2007)
18. Gwizdka, J., Cole, M.: Does interactive search results overview help?: an eye tracking study. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13). ACM, New York, pp. 1869–1874 (2013)