Improving Personal Tagging Consistency through Visualization of Tag Relevancy

Qin Gao[1], Yusen Dai, and Kai Fu

[1] Department of Industrial Engineering, Tsinghua University, Beijing 100084, China
gaoqin@tsinghua.edu.cn

Abstract. Tagging has emerged as a new means of organizing information, but the inconsistency in tagging behaviors of users is a major drawback which degrades both information organization and retrieval performance. The current study aims to study how the intra-personal consistency of tagging can be improved by proper tag visualization. The effects of visualization of tag frequency and visualization of the relevancy among tags on personal tagging consistency are empirically tested and compared through an experiment with 39 participants. The results show that visualization of tag relevancy improves tagging consistency significantly and reduces mental workload simultaneously; visualization of tag frequency may alleviate perceived physical demand when tag relevancy is visualized. The findings provide clear and meaningful implications for system designers.

Keywords: collaborative tagging systems, consistency, tagging, information visualization, tag cloud.

1 Introduction

As web2.0 services gain popularity, previously passive web users are becoming active content creators and organizers. Users are allowed to describe contents using free-chosen keywords, so-called "tags", with the aim to facilitate access for themselves and for other users. The benefits and problems of collaborative tagging systems have been discussed by many researchers. First of all, one of the most obvious advantages of tagging is the very simple rule of tagging. No training or expertise on complicated hierarchically organized nomenclatures is required for using a collaborative tagging system. This simplicity reduces cost of indexing and labeling objects dramatically [1, 2]. In particular, such open structure and flexible usage of collaborative tagging systems make them applicable for domains characterized with large corpus, absence of formal categories, unstable and unrestricted entities where formal taxonomies do not work [3]. Second, in collaborative tagging systems, information-seekers are indexing and labeling information mainly for their own usage. Thus user-generated tags are believed reflecting the mental models of real system users [4]. Third, the bottom-up approach leads to multiple interpretations of the same content, and this enables users to benefit from other people's discovery in addition to their own [5] and encourages both social and cultural explorations [6]. The democratic aspect of collaborative

A.A. Ozok and P. Zaphiris (Eds.): Online Communities, LNCS 5621, pp. 326–335, 2009.

tagging systems makes them more responsive to changes in the users' actual interests [7] and changes in the consensus of how things should be classified [3].

The flexibility of tagging systems, however, also leads to problems that negatively influence the quality of tags and the effectiveness of system use. One important problem is the low level of consistency in tagging. Here consistency refers to two aspects: inter-personal consistency and intra-personal consistency. The former refers to the degree to which different users describe the same content with consistent tags, and the latter refers to the level of consistency in tagging style within individual users. Until now little research has been done to improve the personal consistency. This perspective, however, is indispensable, since it obviously influences the effectiveness and satisfaction of system use by individuals, and consequently the success of the system.

The current study aims to study how the intra-personal consistency of tagging can be improved by proper tag visualization. The effects of visualization of tag frequency and visualization of the relevancy among tags on personal tagging consistency are empirically tested and compared. The results indicate that visualization of tag relevancy improves personal tagging consistency significantly, whereas visualization of tag frequency may alleviate perceived physical demand. The findings provide meaningful implications for system designers.

2 Inconsistencies in Collaborative Tagging Systems and Visualization of Tags

The problem of inconsistencies in collaborative tagging systems has been long discussed. The democracy and freedom in such systems leads to problems of imprecise, overlapping and ambiguous tags. Various "bad" tags exist in the system due to the lack of control and education, including misspelt tags, badly encoded tags, tags that do not follow conventions in issues such as case and number and mixed-use of plurals and singulars [8]. Golder and Huberman discussed those problems related to semantic relations between tags and their referents, including polysemy (one term related to many meanings), synonymy (multiple terms related to the same meanings) and basic level variation (terms at different specificity used arbitrary for one object) [9]. The reason is not difficult to understand. Since users from different cultural and knowledge background add tags with different motivations and goals in different contexts, the whole system is consisted of idiosyncratically personal categories as well as those that are widely agreed upon [9]. Even if all people use similar vocabulary, the determination of the specificity of a tag is influenced by many variable factors, including users' linguistic power of expression, cognitive talents, tagging motivations, domain knowledge and personal living experience [9, 10].

Despite the very idiosyncratic population of users, however, researchers found certain regularities in the usage pattern. Mathes suggested that tag distribution would follow a power law scenario: the most used tags are highly visible so likely to be used by other users, and there will be a large number of tags that are used only by a few users, and finally there will be a huge number of tags that are used by just one or two users [11]. However, by graphing tags with their frequencies of a given URL, it was found that the shape is similar to classic power law, but the drop off of tag frequencies is not as steep as in a power law [12]. Golder and Huberman's analysis of del.icio.us tags detected that after about 100 bookmarks, a stable tagging pattern will arise for a given URL [9]. In addition, Kipp and Campbell found besides those generally agreed tags, other tags follow several inconsistency patterns. However, tags related to time and task reveals a new dimension in information organization, which is not supported by traditional classification schemes.

Tagging consistency, the main dimension of tag quality, as detailed previously in this paper, is a major problem of folksonomy. The extent to which users adhere to a consistent tagging pattern influences not only their usage of the system but also the vocabulary quality of overall system. Generally there are two ways to improve personal tagging consistency: educating users to add better tags and improving the systems to allow better tags to be added [8]. Though fully automated semantic checks by the system sounds attractive, it is nearly impossible due to the ambiguous nature of language and the heterogeneous user population. Providing users with a set of helpful heuristics to facilitate wise tag selection seems a wiser solution, like tag could. Most of currently used tag clouds are developed to visualize the frequency of tags used. The benefit of such visualization is that it encourages the tagger to reuse tags and to direct the tagger to the community consensus. However, there are also concerns that such visualization impedes users to discover semantic relations among tags and reduces efficiency of visual search [13]. Visualization of semantic relations among tags is considered another effective way to improve tagging quality and efficiency [14, 15]. It is expected that such visualization will help users find suitable tags easily, and improve the performance of visual search and navigation. However, there is no empirical evidence supporting such statement yet.

3 Research Questions

Hypothesis B1: visualization of occurrence frequency of tags improves personal tag consistency.

Harry Halpin et al shows that tags distributions in del.icio.us tend to stabilize into power law distributions [16]. This distribution indicates the existence of dominating topics in such a collaborative system. It implies that for a specific user, some topics are more interesting, thus the user tends to tag them more frequently. That is, the frequently used tags are more likely to be reused, and making frequently used tags easy to find in the tag list encourages the user to reuse it. When the user tries to tag a specific object, she may know what tag she will use exactly (then she can type it in directly or search in the tag list, which is relatively easy to accomplish in an alphabetically sorted list), or she just has a brief idea about the tag content. In the latter case, she is performing actually a browsing task. By visualizing the occurrence frequency, we can make the frequently used tags easier to find in such browsing tasks, ultimately, to improve personal consistency.

Hypothesis B2: visualization of inter-tag relevancy improves personal tag consistency.

Lin suggests browsing is facilitated when there is a good underlying structure so that items close to one another can be inferred to be similar[17]. It is expected that visualizing the inter-tag relevancy help users in stabilizing their tagging patterns.

4 Methodology

4.1 Design of the Experiment

A two by two experiment was conducted to test the hypotheses. Independent variables were the visualization of tag frequency and the visualization of tag relevancy. When the visualization of tag frequency was enabled, the tag frequency was visualized by different font size of the tag in the tag-cloud. The font size of tags ranges from 12px to 60px, with seven levels altogether. The definition of font size levels is described in Table 1.

Table 1. Definition of font size levels

Font size level	*Font size (px)*
1	12
2	20
3	28
4	36
5	44
6	52
7	60

For a given tag, the font size was determined by the following logarithm function:

$$Current_i = \left\lceil \frac{6\log(O_i)}{\log(120)} \right\rceil + 1$$

Where $Current_i$ is the font size level of the current tag, O_i is the use frequency of the current tag.

The resulting relationship between the frequency of a tag and its font size level is shown in Fig. 1.

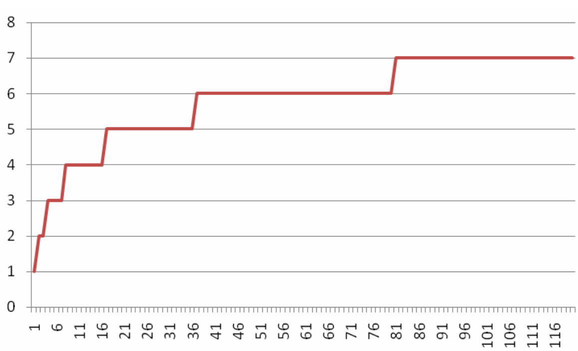

Fig. 1. The relationship between font size level and tag frequency

When the visualization of tag relevancy was enabled, clusters of relevant tags were calculated with K-means algorithm developed by Montero and Solana. By counting the number of co-occurrences, that is, the number of times when two tags are assigned to the same focus, Montero and Solana consider this as a measure of the similarity of different tags, and further as a kind of semantic relationship between the tags [13]. Though the co-existence does not necessarily mean similarity, they may reflect existence of differing user groups and their differing views of the object [12].

To ensure the display suitable for small or large numbers of tags on a web page, we apply following rules on the clustering results

1. Tags of the same cluster are presented in the same line
2. Each line can include at most 10 tags
3. When the cluster has more than 10 members, the first 10 members closest to the central point will be displayed. The rest members will be put to the closest cluster which has room for more members.
4. The number of clusters is [N/6], where N is the number of all available tags.
5. Only clusters with more than 3 members will be displayed.
6. White (RGB value: FFFFFF) and blue (RGB value: 0000FF) background colors are used for adjacent clusters to emphasize the effect of "clusters".

The dependant variable is personal tagging consistency. It is measured by the overlapping between the results of two same tagging tasks [13]:

Let A_i and B_i denote the sets of tags that assigned to the same content i in two successive runs. We define the relative overlapping between A_i and B_i by:

$$O_i(A,B) = \frac{|A_i \cap B_i|}{|A_i \cup B_i|}$$

We define the overall overlapping between the two successive runs by the mean of all relative overlappings of this user:

$$O = \frac{\sum_{i=1}^{n} O_i(A,B)}{n}$$

Fig. 2 shows the condition when both frequency and tag relevancy are visualized.

In addition, the mental workload perceived by participants are also measured using NASA-TLX.

4.2 Participants

43 participants were invited to the tests, yielding 39 valid results, including 10 females and 29 males. Participants were randomly assigned to four groups. The treatment of groups is show in Table 2. Background information of participants is summarized in Table 3.

Fig. 2. Tagging with visualization of both frequency and tag relevancy

Table 2. Treatment of different groups

Group	*Participants*	*Visualization of frequency*	*Visualization of relevancy*
1	10	No	No
2	10	Yes	No
3	9	No	Yes
4	10	Yes	Yes

Table 3. Background information of participants

	Group 1		*Group 2*		*Group 3*		*Group 4*	
Variables	M	SD	M	SD	M	SD	M	SD
Age (years)	23.1	1.29	23.4	3.37	22.2	1.64	22.4	1.17
Computer experience (years)	7.0	2.62	8.9	2.55	7.9	2.75	10.3	2.58
Internet experience (years)	6.1	2.18	7.6	2.05	6.5	1.94	8.1	2.23
Tagging experience (years)	2.25	1.93	2	1.49	2.3	1.22	3.3	1.63

4.3 Procedure

After filling out the consent form and getting familiar with the experiment system, the participant was first asked to tag 60 pictures, in which 20 were experiment stimuli, and 40 were filler pictures. Upon the completion of the tasks, the participant was asked to fill out a NASA-TLX questionnaire, which measured the mental workload of the tagging task utilizing different visualizations. Then the participant was interviewed of how he or she selected tags. After the interview, the participant was asked to do a series of mental arithmetic and spoke the result out loudly. The aim of doing mental arithmetic was to remove participant's memory of tags he or she made in the previous session. Then another 60 pictures were given to the participant to tag, including the identical 20 stimuli they tagged in the previous session and another 40 filler pictures. The sequence of all the pictures was randomized. Finally the participant was briefed and compensated with 50 Yuan RMB.

5 Results and Discussion

5.1 Effect of Relevancy Visualization

The difference in tagging consistency was tested with ANCOVA, with internet experience as the covariate. The result shows that the visualization of tag relevancy improves personal consistency significantly (F =4.37, p = .04). The personal consistency of the relevancy visualization group (M = 0.75, SD = .13) was 12% higher than that of the group without such visualization (M = 0.67, SD = .13).

The comparison of NASA-TLX scores shows no significant difference in the overall score. But regarding mental demand, there is a marginally significant difference (F = 3.0, p = .09). Participants in the relevancy visualization group perceived lower level of mental demand (M = 44.22, SD = 17.56) than participants in the group without such visualization (M = 51.1, SD = 16.47). In addition, there is a trend (F = 2.44, p = .12) that participants in the relevancy visualization group perceived higher level of mental demand (M = 42.2, SD = 21.91) than participants in the group without such visualization (M = 32.7, SD = 18.28).

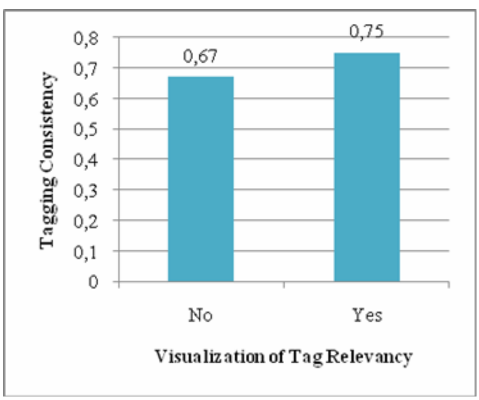

Fig. 3. Effects of relevancy visualization on personal consistency

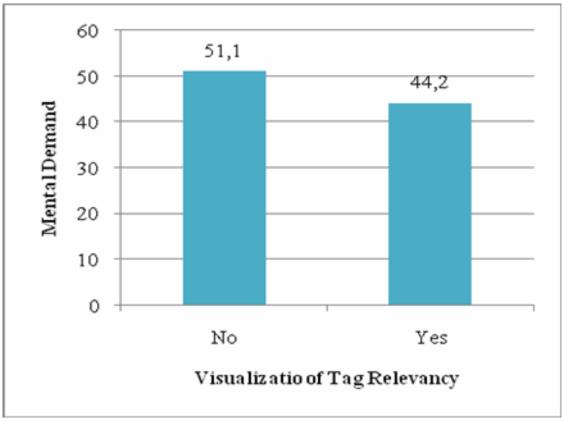

Fig. 4. Effects of relevancy visualization on perceived mental demand

5.2 Effect of Frequency Visualization

The frequency visualization has no significant impacts on tagging consistency and the overall workload. But it influence perceived physical demand significantly (F = 5.2, p = 0.03). Participants in the frequency visualization group perceived lower level of mental demand (M = 22.3, SD = 15.84) than participants in the group without such visualization (M = 33.8, SD = 19.49). More particularly, it is found that there is an interaction effect between frequency visualization and relevancy visualization (F = 4.58, p = .04). As shown in Fig. 5, when the relevancy visualization is enabled, visualization of tag frequency reduces the physical demand for users significantly, whereas such phenomenon does not arise when there is no relevancy visualization.

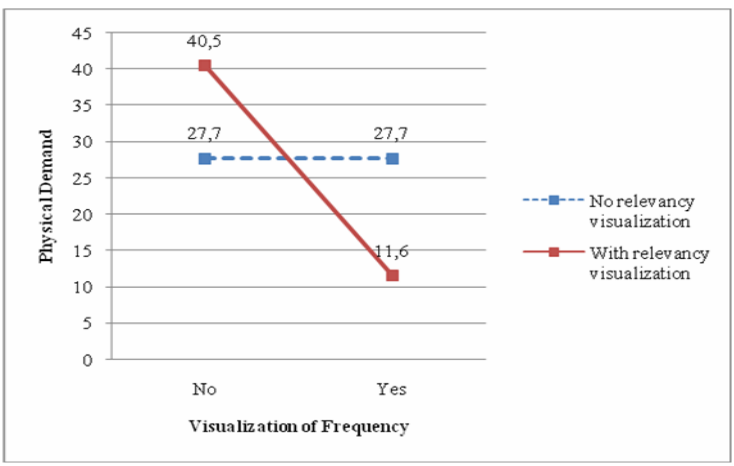

Fig. 5. Effects of frequency visualization on perceived physical demand

6 Conclusion

The experiment results show that relevancy visualization does shape users' tagging behavior into a more consistent pattern, but frequency visualization, though frequently used, does not influence tagging consistency significantly. Interestingly enough, when interviewed after the experiments, most participants in the relevancy visualization reported that they did not realize the semantic clustering of tags.

According to our observation and interviews with participants, we found that there are two different types of tags which exist simultaneously for nearly every participant: categorical tags and descriptive tags. Categorical tags are more general, describing the object from a less ego-centered perspective. The selection of categorical tags is influenced by the so-called basic level [9]. Often such tags are selected at first, and the consistency of such tags is quite high, as the basic level categorization itself is the level of specificity that directly related to humans' interactions with the object. To input such tags, many participants typed directly rather than searching in the tag cloud. The frequency visualization often emphasizes these categorical tags which are already used in a relative consistent pattern. Consequently there is no observed significant difference. The real inconsistency exists in descriptive tags, which describe details of the target in a more ego-centered way. The frequency of such tags is generally low, and they are scattered widely in the tag cloud without relevancy visualization. With the clustering algorithm, however, the co-existence of tags is measured and visualized. Descriptive tags used to describe same targets are placed near to each other. This improves the possibility the user notices relevant descriptive tags and reuses them. In this way, relevancy visualization improves the consistency of descriptive tag usage and the overall tagging consistency. It was also found that visualizing the relevancy among tags also helps reduce mental workload for users.

The results also show that when the relevancy visualization is applied, frequency visualization can alleviate perceived physical demand greatly. This is in consistent with Fitts law, and it provides clear enough implications for designers that frequency visualization should be used in combination of relevancy tags.

In conclusion, we found that frequency visualization and relevancy visualization are both helpful for improving tagging experience, and should be used in combination. There are several alternative algorithms to visualize the relevancy among tags. The current study adopted clustering algorithm based on co-existence of tags. Real impacts of other algorithms should be studied in future research.

References

1. Macgregor, G., McCulloch, E.: Collaborative tagging as a knowledge organisation and resource discovery tool. Library Review 55(5), 291–300 (2006)
2. Wu, H., Zubair, M., Maly, K.: Harvesting social knowledge from folksonomies. In: Proceedings of the 17th Conference on Hypertext and Hypermedia, pp. 111–114. ACM, New York (2006)
3. Shirky, C.: Ontology is Overrated: Categories, Links, and Tags. Economics & Culture, Media & Community (2005)

4. Quintarelli, E.: Folksonomies: Power to the people. In: ISKO Italy UniMIB meeting. Milan (2005)
5. Campbell, K.E.: A phenomenological framework for the relationship between the semantic Web and user-centered tagging systems. In: 17th ASIS&T SIG/CR Classification Research Workshop (2006)
6. Auray, N.: Folksonomy: the New Way to Serendipity. International Journal of Digital Economics 65, 67–88 (2007)
7. Niwa, S., Doi, T., Honiden, S.: Web page recommender system based on folksonomy mining for ITNG 2006 submissions. In: Third International Conference on Information Technology: New Generations, ITNG 2006 (2006)
8. Guy, M., Tonkin, E.: Folksonomies: Tidying up tags? D-Lib Magazine 12 (2006)
9. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science 32(2), 198–208 (2006)
10. Stock, W.G., Peters, I.: Folksonomy and information retrieval. In: Proceedings of the 70th ASIS&T Annual Meeting, vol. 44, pp. 1510–1542 (2007)
11. Mathes, A.: Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Graduate School of Library and Information Science, University of Illinois Urbana-Champaign (2004)
12. Kipp, M., Campbell, D.: Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. Proceedings of the American Society for Information Science and Technology 43(1) (2006)
13. Hassan-Montero, Y., Herrero-Solana, V.: Improving tag-clouds as visual informatin retreival interfaces. In: The International Conference on Multidisciplinary Information Sciences & Technologies (2006)
14. Choy, S., Lui, A.: Web Information Retrieval in Collaborative Tagging Systems. IEEE Computer Society, Washington (2006)
15. Nielsen, M.: Functionality in a second generation tag cloud Nielsen, Mogen, 1980-Functionality in a second generation tag cloud, Gj?vik (1980)
16. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: WWW 2007, Banff, Alberta, Canada (2007)
17. Lin, X.: Map displays for information retrieval. Journal of the American Society for Information Science 48(1) (1997)