# Dynamic User Tag Cloud Construction in Social Tagging (Postprint)

**Authors:** Xie Mengyao, Xuwei Pan

**Date:** 2017-11-08T00:00:00+00:00

## Abstract

[Objective] Tag clouds can be utilized for information retrieval, recommendation, and navigation. Since user tagging exhibits temporal characteristics, to effectively reveal the dynamic changes in user interests, we propose a method for constructing user dynamic tag clouds based on temporal evolution. [Method] By leveraging the forgetting and reinforcement characteristics of memory from psychology, we construct dynamic weights for tags, thereby establishing user dynamic tag clouds that reflect changes in user focus. [Results] Compared with existing tag cloud algorithms, the proposed user dynamic tag cloud algorithm can effectively rank tags according to users' dynamically changing interests, demonstrates significantly superior performance in predicting user interest tags, and achieves higher recommendation accuracy. [Limitations] Since user interests do not change substantially within short time periods, the prediction performance of the dynamic method is not very pronounced in short-term cycles, but is more significant over longer time periods. Conclusion The user dynamic tag cloud based on temporal evolution can effectively capture users' current interest hotspots and improve the effectiveness of personalized retrieval and navigation.

## Full Text

## Constructing Dynamic User Tag Clouds in Social Tagging

**Xie Mengyao, Pan Xuwei**
School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou 310018, China

### Abstract

**[Objective]** Tag clouds serve as effective tools for information retrieval, recommendation, and navigation. Since user tagging exhibits temporal characteristics, this paper proposes a method for constructing dynamic user tag clouds based on temporal evolution to effectively reveal the dynamic changes in user interests. [**Methods**] We establish dynamic weights for tags by leveraging the forgetting and strengthening characteristics of memory from psychology, thereby creating dynamic user tag clouds that reflect shifts in user attention. [**Results**] Compared with existing tag cloud algorithms, the proposed dynamic tag cloud algorithm can effectively rank tags according to dynamically changing user interests, demonstrating significantly higher predictive performance for user interest tags and achieving greater recommendation accuracy. [**Limitations**] Because user interests do not change substantially within short time periods, the dynamic method' s predictive effects are less pronounced over short cycles but become more significant over longer periods. [**Conclusions**] The dynamic tag cloud based on temporal evolution can effectively capture users' current interest hotspots and improve the effectiveness of personalized retrieval and navigation.

**Keywords:** Social Tagging; Tag; User Interests; Dynamic Tag Cloud
**Classification:** TP311

### Introduction

Web 2.0 has brought substantive changes to the Internet, transforming user roles from passive recipients of online information to active creators. As a core component of Web 2.0, social tagging allows mass users to unconstrainedly annotate resources of interest based on their own understanding, with all user annotations being mutually visible [1]. Social tagging sites centered on mass user participation, represented by Flickr and Delicious, have gradually grown and developed, becoming a new channel for users to effectively access information resources.

Tags, as the carrier of social tagging systems, have become an important information organization tool in the Web 2.0 era. Social tags contain rich information, and tag clouds—collections of tags with visual weight—have effectively solved the visualization problem of tag information, helping users quickly obtain valuable information from large numbers of tags. Current research on tag clouds primarily focuses on sorting algorithms [2], personalized recommendations [3], and visual layout [4-5], which has strongly promoted the development of tag cloud applications and theory. As an emerging retrieval and recommendation technology, tag clouds guide user browsing by visually representing the importance of different tags, thereby directing user attention to specific fields or regions. Millen et al. [6] studied user query and browsing habits in tag cloud environments, finding that social tags are an important way to improve social navigation. Hassan-Montero et al. [5] calculated tag usefulness by defining metrics such as the degree to which tags describe resources and the number of resources covered, and improved browsing experience through clustering algorithms. Additionally, as a navigation interface for social information, tag clouds can provide personalized search recommendations to different users after visualizing tag attributes and content classification. Xia et al. [2] processed information through tag cloud construction based on the structure and content of the Wikidata knowledge base, ultimately achieving information retrieval and page ranking.

Existing tag clouds primarily calculate tag weights quantitatively based on cumulative annotation frequency, using different colors or font sizes for intuitive visualization to facilitate user retrieval and browsing. Tags used by users can reflect their interests to a certain extent. However, as time passes, user interest preferences and focus points change, and existing tag clouds built on cumulative frequency cannot effectively reflect these changes. Therefore, how to construct dynamic user tag clouds based on tags used at different times to reveal changes in user interests and focus has become an important issue in leveraging tag clouds to better support user information retrieval and navigation. To address this, this paper investigates methods for constructing tag clouds that reflect dynamic changes in user interests from the perspective of temporal characteristics in user tagging.

## 2 Constructing Dynamic User Tag Clouds

Constructing dynamic user tag clouds requires full consideration of the impact of temporal information on user tag usage. Based on the forgetting characteristics and memory strengthening phenomena in psychology, we view the entire user tagging process as a combination of forgetting and repeated learning processes. Specifically, tags annotated farther in the past have lower importance, while repeatedly appearing tags have strengthened importance. Through this process, we dynamically calculate the weights of different tags to construct dynamic tag clouds and improve user browsing experience.

**2.1 Dynamic Update of Tag Weights** User interests continuously change over time, and this change represents a forgetting phenomenon. According to the forgetting characteristics and memory strengthening phenomena in psychology [7], the following basic features exist: (1) interests closer to the current moment have higher weights, and interest weights gradually decline over time; (2) when the same interest repeatedly appears, there is a reinforcement process that merges with the original interest to form new user interests. Therefore, each interest undergoes both forgetting and repeated learning processes. Tags are expressions of users' own attitudes and interests, so we can utilize the forgetting and strengthening characteristics of interest memory to update the dynamic weights of user tags to reflect the impact of time on tag weights, thereby supporting the construction of dynamic tag clouds.

Using weights to measure users' interest level in tag $t_k$, the tag weight $w_k$ undergoes decay and strengthening processes, forming a multi-stage decay process as shown in [Figure 1: see original paper]. In [Figure 1: see original paper], during a certain time period (e.g., from $d_{n0}$ to $d_{n1}$), the tag weight $w_k$ decays over time. When users continuously annotate in social tagging systems, the same interest periodically reappears (e.g., tag $t_k$ reappears at times $d_{n1}$ and $d_{n2}$), and $w_k$ is strengthened and rises again. Such repeated activities divide the entire user annotation process into multiple sub-stages, each representing a new forgetting process. Therefore, based on similar forgetting curve formulas constructed by Yu et al. [7] and Yin et al. [8], we improve upon them and propose a formula for calculating dynamic tag weights. The calculation of tag $t_k$' s dynamic weight $w_k$ involves three main components: weight calculation at specific time points, forgetting decay, and memory strengthening.

**(1) Tag Weight Calculation at Time Points**

The weight $w$ of tag $t_k$ at a specific time point $T$ is calculated using the TF (term frequency) method, i.e., the proportion of times tag $t_k$ is used to the total number of tag uses at that time point (e.g., a specific day). The calculation method is shown in formula (1):

$$w = \frac{f_k}{\sum_{s=1}^{m} f_s} \tag{1}$$

where $m$ is the total number of tags at that time point, and $f_s$ is the frequency of tag $t_s$.

**(2) Tag Weight Forgetting Decay**

If tag $t_k$ does not reappear, its weight $w_k$ decays over time, which can be calculated using an exponential forgetting function. The quantitative function for the forgetting process is defined as formula (2):

$$w_k^d = w_k^{d_{n-1}} \times 2^{-\frac{d-d_{n-1}}{hl_u}} \tag{2}$$

where $w_k^d$ is the decayed tag weight, $w_k^{d_{n-1}}$ is the weight when tag $t_k$ appeared for the $(n-1)$th time (i.e., the initial value of the previous forgetting stage); $hl_u$ is the half-life of user $u$, which varies with the user' s knowledge acquisition behavior cycle; and $d - d_{n-1}$ represents the time difference since tag $t_k$ last appeared.

**(3) Tag Weight Memory Strengthening**

As shown in [Figure 1: see original paper], at three time points $d_{n1}$, $d_{n2}$, and $d_{n3}$, tag $t_k$ repeatedly appears. The value of $w_k^{d'}$ consists of the remaining weight from the previous stage's decay and the weight increase from the new annotation activity of the same tag $t_k$. Formula (3) calculates the initial interest degree for each forgetting stage:

$$w_k^{d'} = w_k^{d_{n-1}} \times 2^{-\frac{d_n-d_{n-1}}{hl_u}} + w_k^{d_n} \tag{3}$$

where $w_k^{d'}$ is the initial tag weight at time point $d_n$, $d_n$ represents the time point when tag $t_k$ appears for the $n$th time, so $d_n - d_{n-1}$ is the time difference between two consecutive appearances of tag $t_k$; $w_k^{d_n}$ is the weight when tag $t_k$ appears for the $n$th time, whose calculation method is given by formula (1) and represents the weight increase from annotation activity at time point $d_n$; the first term represents the remaining weight from the previous stage' s decay to time point $d_n$.

**2.2 Dynamic Tag Cloud Construction Algorithm** Based on the above dynamic tag weight update mechanism, we establish the following dynamic tag cloud construction algorithm.

**Input:** User $u$' s annotation history records (including annotation time, resources, and tags used)
**Output:** Dynamic tag cloud for user $u$

**Algorithm Description:** 1. Use formula (1) to calculate the initial tag interest weights from user annotations, obtaining each tag' s weight at different time points (typically calculated daily). 2. Sort tags in chronological order and determine whether tag $t_k$ reappears. If not, proceed to step 3 to update the tag weight; if it does, proceed to step 4 to update the tag weight. 3. Calculate the decayed tag weight using formula (2). 4. Calculate the strengthened tag weight using formula (3), which combines the decayed value from the previous stage and the weight increase from the new annotation activity. 5. Synthesize each tag' s weight and perform normalization to obtain user $u$' s dynamic tag cloud.

## 3 Experimental Evaluation

**3.1 Experimental Data** The experimental data comes from two representative social tagging systems: Last.fm and Delicious. The Delicious data was collected by the DAIM research group at Peking University, covering over 185,000 users' social tagging data from Delicious website between January and June 2009, available at http://www.datatang.com/data/42989. The Last.fm data was collected by the Information Retrieval Group at Universidad Autónoma de Madrid, containing music annotation data from 1,892 users, available at http://grouplens.org/datasets/hetrec-2011. The basic statistics of the experimental data are shown in .

Basic Statistics of Experimental Data

**3.2 Visualization of Dynamic User Tag Clouds** To visually reflect the effect of dynamic tag clouds, we compare the visualization results of typical users' tag clouds. We construct visual tag clouds using both the existing cumulative annotation frequency method and the proposed dynamic tag cloud construction method, with font size distinguishing tag weights. [Figure 2: see original paper] shows the visualization results of tag clouds constructed by the two methods for an active Delicious user (UserID: 12116) at the end of the 6th month (i.e., annotation cutoff time), displaying the top 50 popular tags, where larger font size indicates higher tag weight. This user performed 2,710 annotations within 6 months, covering 995 resources and using 424 tags, with the most frequently used tag appearing 447 times and the least frequent appearing once.

Example Annotation Data

**3.3 Evaluation of Dynamic Tag Cloud Navigation Effectiveness** To further verify whether the proposed dynamic tag clouds better reflect changes in user interests and provide better information retrieval and navigation, we conducted quantitative comparative experiments. The underlying assumption is that if a user is currently interested in a certain tag, they will continue to use it in the future. Therefore, we construct evaluation metrics based on whether top-N tags with high weights at a certain time point are used again by the user in a subsequent period, to characterize how well the tag cloud captures user interests. We first define two basic evaluation metrics as shown in formulas (4) and (5):

$$Acc = \frac{N_n}{N} \tag{4}$$

$$Rec = \frac{N_n}{N_0} \tag{5}$$

where $N_n$ is the total number of times the top-N tags are used in the future period, and $N_0$ is the total number of times all tags are used by the user in that future period. As can be seen from the definitions, $Acc$ represents the average frequency of top-N tags used in the future period, while $Rec$ represents the ratio of total usage frequency of top-N tags to all tags used by the user during that period. These two basic metrics characterize the usage of top-N tags from different perspectives. We therefore define a new comprehensive evaluation metric combining both, as shown in formula (6):

$$AR = Acc \times Rec \tag{6}$$

Taking the Delicious dataset as an example, we select 15 users with relatively complete annotation histories (i.e., those with basically 6 months of continuous annotation activity) as subjects. The 6-month period is divided into 36 chronological cycles, each lasting 5 days. For $k = 1, 2, \ldots, 36$, we take the tag weights at the end of the $k$th cycle, rank the top 10, 20, and 30 tags, calculate their usage in the next cycle' s 5-day period, and compute the $AR$ evaluation metric. For example, when $k = 2$, we calculate the top 10, 20, and 30 tags by weight up to day 10, then use the user' s annotation data from the next 5 days (i.e., the 3rd cycle from day 11 to day 15) to calculate the metric values. To avoid the impact of randomness in the initial annotation stage, we start calculating the corresponding metric values from the 20th cycle. Similarly, for Last.fm, we select 15 active users as research subjects. Given the large time span of this dataset, we divide the 6-year period into 12 cycles of 6 months each, using the same evaluation method as for Delicious.

To examine the difference between the dynamic tag clouds constructed using temporal annotation information and static tag clouds built on cumulative annotation frequency in reflecting user interests, we calculate their $AR$ values and compute their ratio as shown in formula (7):

$$Ratio = \frac{AR_D}{AR_S} \tag{7}$$

where $AR_D$ and $AR_S$ are the comprehensive evaluation metric values calculated by the dynamic and static tag cloud methods, respectively. [Figure 3: see original paper] shows the comparison results for 15 users using both methods. It is evident that for the top 10, 20, and 30 tags, the dynamic method shows varying degrees of improvement over the static method, with improvements ranging from 4% to 11% on the Delicious dataset and more significant improvements of 18% to 82% on the Last.fm dataset.

As seen in the visual tag clouds in [Figure 2: see original paper], the static tag clouds constructed using existing cumulative annotation frequency methods and the dynamic tag clouds constructed using the proposed temporal annotation information method produce different relative tag weights, thus generating different effects in guiding users' relative priority for information retrieval and navigation.

Furthermore, under the same Last.fm experimental data and evaluation metrics, we conducted comparative experiments between the dynamic method and other approaches, selecting the cumulative frequency method (TF) and the tag time-weight strategies proposed in literature [8], including TF time-weight and TFIDF time-weight. As shown in [Figure 4: see original paper], the dynamic method outperforms other methods with higher recommendation accuracy.

[Figure 4: see original paper] AR Values of Four Different Methods

### Conclusion

To enable tag clouds to better reflect users' current dynamic interests, this paper proposes a dynamic user tag cloud construction method based on temporal evolution, leveraging the characteristics of dynamic user interest changes and the temporal features of social tagging. This method builds dynamic tag weights based on the forgetting and strengthening characteristics of memory in psychology. Experimental results show that the visualization of dynamic tag clouds differs from static tag clouds constructed using cumulative annotation frequency. Compared with existing tag sorting algorithms, the dynamic method is superior and can effectively capture and grasp users' current interests, facilitating better information retrieval and navigation through tag clouds. Since tags are only single words reflecting user interests, while user interests are often characterized by themes formed by multiple tags, further mining of user interest themes based on the current dynamic tag clouds will be the focus of future work.

## References

[1] Wei Jianliang, Zhu Qinghua. Advances in Personalized Information Recommendation Based on Social Tagging[J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(4): 625-633.

[2] Xia Yu, Ge Jiaqi, Ma Xiu, et al. Research on Searching Algorithm Based on Wikidata and Tag Cloud[J]. Software Guide, 2016, 15(8): 42-46.

[3] Zeng Ziming, Zhang Zhen. A Personalized Recommendation Approach Based on Community Tag Cloud in Social Tagging System[J]. Journal of Intelligence, 2011, 30(10): 128-133.

[4] Bi Qiang, Zhou Shanshan, Ma Zhiqiang, et al. Study on Optimization Mechanism of Tag Cloud for Knowledge Relation[J]. New Technology of Library and Information Service, 2014(5): 33-40.

[5] Hassan-Montero Y, Herrero-Solana V. Improving Tag-Clouds as Visual Information Retrieval Interfaces[C]//Proceedings of International Conference on Multidisciplinary Information Sciences and Technologies, Merída, Spain. 2006.

[6] Millen D R, Feinberg J. Using Social Tagging to Improve Social Navigation[OL]. [2016-09-30]. https://www.researchgate.net/publication/228904554_{{{Using}}*{{social}}*}{{{taggin

[7] Yu Hongtao, Cui Ruifei, Dong Qinqin. Micro-blog User Interest Model Based on Forgetting Curve[J]. Computer Engineering and Design, 2014, 35(10): 3367-3372, 3379.

[8] Yin Guisheng, Cui Xiaohui, Ma Zhiqiang. Forgetting Curve-based Collaborative Filtering Recommendation Model[J]. Journal of Harbin Engineering University, 2012, 33(1): 85-90.

[9] Zhao Kaihui. Recommendation Method of Personalized Information Based on Socialized Tagging[J]. Information Science, 2015, 33(6): 39-42.

## Author Contributions

Xie Mengyao: Methodology research and data experiments, manuscript drafting.
Pan Xuwei: Research idea proposal, experimental design, and manuscript revision.

## Conflict of Interest Statement

All authors declare that they have no conflict of interest.

## Supporting Data

Supporting data [1] is available in the online version of the journal at http://www.infotech.ac.cn. Supporting data [2-4] are self-archived by the authors and available via E-mail: panxw@zstu.edu.cn.

[1] Xie Mengyao, Pan Xuwei. Experimental Data.zip. Selected user experimental data from Delicious and Last.fm.
[2] Xie Mengyao, Pan Xuwei. Static vs Dynamic Method Comparison.xlsx. Improvement effects of dynamic method over static method.
[3] Xie Mengyao, Pan Xuwei. Dynamic Method vs Other Time-Weight Methods Comparison.xlsx. Improvement effects of dynamic method over other methods.
[4] Xie Mengyao, Pan Xuwei. Experimental Database File.sql. Database code.

**Received:** 2016-10-12
**Revised:** 2016-12-09

*Note: Figure translations are in progress. See original paper for figures.*