Analysis of Music Tagging and Listening Patterns: Do Tags Really Function as Retrieval Aids?

Jared Lorince[1] (✉), Kenneth Joseph[2], and Peter M. Todd[1]

[1] Cognitive Science Program Indiana University, Bloomington, Indiana, USA
{jlorince,pmtodd}@indiana.edu
[2] Computation, Organization, and Society Program Carnegie Mellon University, Pittsburgh, PA, USA
kjoseph@cs.cmu.edu

Abstract. In collaborative tagging systems, it is generally assumed that users assign tags to facilitate retrieval of content at a later time. There is, however, little behavioral evidence that tags actually serve this purpose. Using a large-scale dataset from the social music website Last.fm, we explore how patterns of music tagging and subsequent listening interact to determine if there exist measurable signals of tags functioning as retrieval aids. Specifically, we describe our methods for testing if the assignment of a tag tends to lead to an increase in listening behavior. Results suggest that tagging, on average, leads to only very small increases in listening rates, and overall the data do *not* support the assumption that tags generally serve as retrieval aids.

Keywords: Collaborative tagging · Folksonomy · Music listening · Memory cues · Retrieval aids · Personal information management

1 Introduction

In social tagging systems, users assign freeform textual labels to digital content (music, photos, web bookmarks, etc.). There are a variety of reasons for which users tag content, but it is overwhelmingly assumed that tagging for one's own future retrieval – assigning a tag to an item to facilitate re-finding it at a later time – is users' principal motivator. But is this a valid assumption?

Collaborative tagging systems are often designed, at least in part, as resource management platforms that expressly facilitate the use of tags as retrieval aids. However, the freeform, and often social, nature of tagging opens up many other possible reasons for which a user might tag a resource. While there is a significant amount of non-controversial evidence for such alternative tagging motivations (sharing resources with other users, social opinion expression, etc.), the problem with the retrieval aid assumption runs deeper than there simply existing possible

N. Agarwal et al. (Eds.): SBP 2015, LNCS 9021, pp. 141–152, 2015.
DOI: 10.1007/978-3-319-16268-3_15

alternatives. There is, in fact, almost no behavioral evidence that tags are ever actually used as retrieval aids. While there is much data available on user tagging habits (i.e. which terms are applied to which resources, and when), to our knowledge there is no published research providing behavioral evidence of whether or not tags, once applied to items, actually facilitate subsequent retrieval. This is an issue largely driven by a lack of data: Although a web service can in principle track a users' interaction with tags (for instance, if users use tags as search terms to find tagged content), there are no available datasets containing such information, nor can it be crawled externally by researchers.

Despite these issues, this empirical question is not intractable. While detailed information on how existing tags are utilized remains beyond our reach, an alternative approach is to examine how patterns of user interaction with tagged versus untagged content vary. If tags do serve as retrieval aids, we should expect users to be more likely to interact with a resource (e.g. visit bookmarked pages, listen to songs, view photos, etc.) once they have assigned a tag to it.

Here we test this hypothesis using a large-scale dataset consisting of complete listening and tagging histories from more than 100,000 users from the social music website Last.fm. From this dataset, we extract user-artist listening time-series, each of which represents the frequency of listening over 90 months to a particular artist by a particular user, and compare time-series in which the user has tagged the artist to those that are untagged. Specifically, we address the following two questions:

RQ1: Does tagging an artist lead to increased listening to that artist in the future, as shown by comparison of tagged versus untagged time-series?

RQ2: Are certain tags particularly associated with increases in future listening, and if so, can we identify attributes of such "retrieval-targeted" tags as opposed to others?

2 Background

Collaborative tagging has been considered one of the core technologies of "Web 2.0", and has been implemented for resources as diverse as web bookmarks (Delicious), photos (Flickr), books (LibraryThing), academic papers (Mendeley), and more. Vander Wal [15] coined the term "folksonomy" to describe the emergent semantic structure defined by the aggregation of many individual users' tagging decisions in such a system. These folksonomies have since become the target of much academic research. One of the earliest analyses of a collaborative tagging system is Golder and Huberman's [4] work on the evolution of tagging on Delicious.com. Around the same time, Hotho and colleagues [9] presented a formal definition of a folksonomy: $\mathbb{F} := (U, T, R, Y)$ [1]. The variables U, T, and R represent, respectively, the sets of users, tags, and resources in a tagging system, while Y is a ternary relation between them ($Y \subseteq U \times T \times R$).

Since 2006, an extensive literature on *how* people tag has also developed, covering topics like tagging expertise [17], mathematical [2] and multi-agent [11]

[1] This is a slight simplification. For details, see [9]

models of tagging choices, consensus in collaborative tagging [6], and much more. Our understanding of the dynamics of tagging behavior has greatly expanded, but the question of *why* people tag has proven much more elusive.

It is typically assumed that tags serve as retrieval aids, allowing users to re-find content to which they have applied a given tag (e.g. a user could click on or search for the tag "rock" to retrieve the songs she has previously tagged with that term). This assumption is central to Vander Wal's original defintion of a folksonomy, "the result of personal free tagging of information and objects... *for one's own retrieval*" [15, emphasis added]. This perspective is echoed in many studies of tagging patterns [3,4,6].

But while retrieval is the most commonly assumed motivation for tagging, other reasons certainly exist, and various researchers have developed taxonomies of tagging motivation. Among proposed motivational factors in tagging are personal information management (including but not limited to tagging for future retrieval), resource sharing, opinion expression, performance, and activism [1,8,18] – see [5] for a review.

While the development of motivational theories in tagging is useful, there is almost no work actually grounding them in behavioral observations. The vast majority of existing work either makes inferences about motivation based on design features of a website (e.g. social motivations in tagging require that one's tags be visible to other users, [13]), employs semantic analysis and categorization of tags (e.g. the tags "to read", "classical", and "love" can be inferred to have different uses, [18]), or directly asks users why they tag using survey methods [1]. The results of such approaches are useful contributions to the field, but few have resulted in testable behavioral hypotheses that can confirm or refute their validity.

One notable exception is work by Körner and colleagues [10]. They argue that taggers can be classified on a motivational spectrum from categorizers (who use a constrained vocabulary suitable to future browsing of their own tagged resources) to describers (who use a large, varied vocabulary to facilitate future keyword-based search with their own tags), and have developed and tested quantifiable signals of these different motivations. The main deficiency of this approach, however, is that their hypotheses are based fully on attributes of user tag vocabularies; they present no way to test whether or not describers actually use tags, once applied, for keyword-based search and categorizers use them for browsing.

Again, the problem of lack of verification arises because data on how users actually *use* existing tags is simply not available to researchers through any tagging system APIs (or through other methods) that we are aware of. Thus the existing work on tagging motivation is limited to inferring *why* people tag from *how* they tag, rather than from how they *use* their tags. In presenting our novel methods, we are aware that they still represent an inferential approach. Our approach is distinct from those just described, however, in that we test a concrete hypothesis about how tagging should affect a behavior on which we *do* have data: interaction with tagged content, in our case music listening.

3 Dataset

Last.fm incorporates two specific features that are of interest for our analysis. First, it implements a collaborative tagging system (a "broad" folksonomy, following Vander Wal's [14] terminology, meaning that multiple users tag the same, publicly available content) in which users can label artists, albums, and songs. Second, the service tracks users' listening habits both on the website itself and on media players (e.g. iTunes) via a software plugin. This tracking process is known as "scrobbling", and each timestamped instance of a user listening to a particular song is termed a "scrobble".

Here we use an expanded version of a dataset described in earlier work [11,12] that includes the full tagging histories of approximately 1.9 million Last.fm users, and full listening histories from a subset of those users (approximately 100,000) for a 90-month time window (July 2005 - December 2012, inclusive). Data were collected via a combination of the Last.fm API and direct scraping of publicly available user profile pages. For further details of the crawling process, see [11,12].

For our current purposes, we consider only those users for whom we have both tagging and listening histories. For each user, we extract one time-series for each unique artist listened to by that user. Each user-artist listening time-series consists of a given user's monthly listening frequency to a particular artist for each month in our data collection period, represented as a 90-element vector (each element of the vector represents the number of times that particular user listened to that particular artist in the particular month).

We selected a monthly timescale for listening behavior due to the fact that user tagging histories are only available at monthly time resolution. Furthermore, we perform all analyses here at the level of artists, rather than individual songs. Thus every song scrobbled is treated as a listen to the corresponding artist, and all annotations (i.e., particular instances of applying a tag to a song, album, or artist) are treated as annotations of the corresponding artist. Our choice to perform all analyses at the level of artists, rather than individual songs, is based on the facts that (a) listening and tagging data for any particular song tend to be very sparse, and (b) the number of time-series resulting from considering each unique song listened to by each user would be prohibitively large.

The 2-billion-plus individual scrobbles in our dataset generate a total of approximately 95 million user-artist listening time-series. In about 6 million of these cases, the user has assigned at least one tag to the artist (or to a song or album by that artist) within the collection period (we refer to these as tagged time-series), while in the remaining cases (approx. 89 million) the user has never tagged the artist. We summarize these high-level dataset statistics in Table 1. Comparison of these tagged and untagged listening time-series is the heart of the analyses presented in the next section.

Table 1. Dataset summary

Total users	104,829
Total scrobbles	2,089,473,214
Unique artists listened	4,444,119
Unique artists tagged	1,049,263
Total user-artist listening time-series	94,875,106
Total tagged time-series	5,930,594
Total untagged time-series	88,944,512

4 Analyses and Results

4.1 RQ1: Comparison of Tagged and untagged Time-Series

Our principal research question is whether listening patterns for tagged content are consistent with the expectation that tags serve as memory cues. If so, we would expect to see an increase in a user's listening rate to musical artists after the user has tagged them, under the assumption that a tag facilitates retrieval and increases the chances of a user listening to a tagged artist.

Unfortunately, several factors combine to make such an analysis difficult. First and foremost, the desired counterfactual of the untagged "version" of a particular tagged series, which would allow a direct testing of how tagging changes listening behavior, does not, of course, exist. We thus must utilize untagged time-series in a way that allows them to approximate what a true counterfactual might look like. In searching for such samples, a second difficulty that arises is that listening rates for tagged time-series are much greater than for untagged time-series (the average number of total listens across time-series is 16.9 when untagged and 98.9 when tagged). While suggestive of the importance of tagging, this imbalance also suggests that controls must be incorporated in both sample selection and statistical analysis to account for previous listening behavior prior to tagging. Finally, the actual point in time at which tags are expected to increase listening behavior for any given user is unknown. Thus, we must formulate our analysis to account for this uncertainty.

To alleviate issues with the non-existence of a true counterfactual, we subselect from both the tagged and untagged series using the following formal procedure. We first select only those tagged time-series that have:

- more than 25 total listens;
- a peak in listening at least 6 months from the edges of our data collection period (ensuring that the period from 6 months before to 6 months after the peak does not extend beyond the limits of our data range); and
- at least one listen in the 6 months prior to and after the peak (e.g. if the peak occurs in July, there should be at least one listen between January and June, and one between August and the following January).

We then select only those tagged time series where the tag was applied in the month of peak listening, and align all of those series at that peak point.

Constraining our time-series in this manner, we are left with a total of 206,140 tagged time-series. Next we randomly select a same-sized sample of the 4.1M untagged time-series meeting the same criteria, and also align them at the peak of listening. Where the peak was reached in multiple months in any series, we chose one of these peaks at random to align on. All results below have been verified with multiple random samplings of the untagged data.[2]

After temporally aligning the tagged and untagged samples, we limited our analysis to a 13-month period extending from 6 months prior to the peak month to 6 months after the peak. This allows us to consider a manageable variety of ways in which listening prior to the tag may affect future behavior.

In Figure 1a we plot mean listens, with 95% normal confidence intervals, for each month across all tagged and untagged time-series in the subsampled data. All values are normalized by the peak number of listens for each series, and thus values at the peak month for both the tagged and untagged lines are unity. Comparing the line heights before and after the peak, Figure 1a shows that the mean normalized listening rate increases in the months after the peak for both tagged and untagged time-series. But there is also a small but reliable difference between the tagged and untagged series: The tagged time-series show proportionally higher mean normalized listening rates after the peak month (in which the the tag was applied) as compared to untagged time-series. This is suggestive of an increase in listening as a result of tagging.[3]

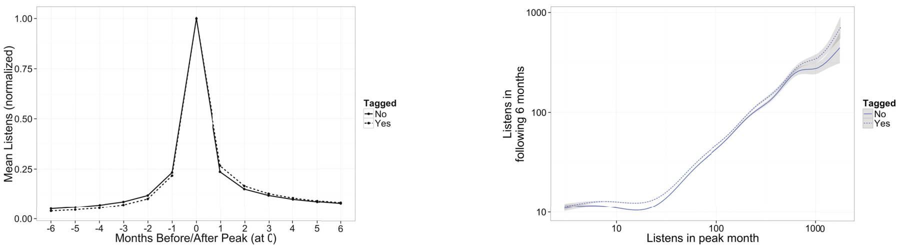
While Figure 1a thus supports our hypothesis, there are two important caveats. First, as the distribution of the number of listens in any given month across all time series is heavily skewed, the mean is not fully representative of the data. Our further statistical analysis uses a log-transformed version of the listening counts to account for this. Second, the initial analysis does not control for the presumably important effect of pre-peak listening behavior on post-peak listening.

To test our hypothesis more robustly, we therefore use a regression model that incorporates previous listening behavior to predict post-peak behavior. Due to a lack of knowledge about the relationship between these variables and the volume of data we have, it was both unreasonable and unnecessary to assume a linear relationship between the dependent and independent variables. Because of this, we opted for a Generalized Additive Model (GAM, [7]) using the R package mgcv [16] applied to all of the tagged and untagged series in the selected sets described above. Our dependent variable in the regression is the logarithm of the sum of all listens in the six months after a tag has been applied, to capture the possible

---

[2] Our method thus compares tagged and untagged data aggregated over many users. While it would be preferable to perform a within-subjects analysis (i.e. comparing tagged versus untagged data for each individual user), thereby accounting for much of the variability in listening across different users, the data for any particular user tends to be too sparse, as most individuals have tagged only a few times (if ever).

[3] There is also, however, a small but reliable lower rate of listening to tagged artists versus untagged artists prior to peak listening. This may indicate that songs that "catch on" for a user more quickly (rise faster in listening from before the peak to the peak) are more likely to be tagged, a possibility to be explored in future work.



(a) (b)

Fig. 1. Comparison of tagged and untagged listening time-series. Mean normalized listens by month (a), and regression results (b), both with 95% confidence intervals.

effect of tagging over a wide temporal window. (The results are qualitatively the same when testing listening for each individual month.) Our independent variables are a binary indicator of whether or not the time-series has been tagged, as well seven continuous-valued predictors, one each for the logarithm of total listens in the peak month and the six previous months. The regression equation is as follows, where m corresponds to the month of peak listening, L is the number of listens in any given month, T is the binary tagged/untagged indicator, and f represents the exponential-family functions calculated in the GAM (there is a unique function f for each pre-peak month):

$$\log \sum_{i=1}^{6} L_{m+i} = b_0 + b_1 T + \sum_{i=0}^{6} f(\log L_{m-i}) \tag{1}$$

The regression model, which explained approximately 30% of the variance in the data (adjusted R^2), indicated that the tagged/untagged indicator and the listening rate parameters (smoothed using thin-plate regression splines) for all seven previous months had a significant effect on post-peak listening behavior ($P \ll 0.0001$). As we cannot show the form of this effect for all model variables at once, Figure 1b instead displays the predicted difference in listening corresponding to tagging as a function of the number of peak listens, calculated with a similar model which considers only the effect of listening in the peak month on post-peak listening. This plot suggests and the full model confirms that, controlling for all previous listening behavior, a tag increases the logarithm of post-peak listens by .147 (95% CI = [.144,.150]). In other words, the effect of a tag is associated with around 1.15 more listens over six months, on average, than if it were not to have been applied. The large confidence interval on the right hand side of Figure 1b reflects the small number of userswho have extremely high listening rates for particular artists.

4.2 RQ2: Tag Analysis

To examine if and how different tags are associated with increased future listening, we ran a regression analysis similar to that described above, but with two important changes. First, instead of a single tagged/untagged indicator, we included binary (present/not present) regressors for the 2,290 unique tags that had at least five occurrences in our subsample.[4] Second, due to the data-hungry nature of the GAM and the large number of additional variables introduced by utilizing all tags as unique predictors, we chose to only control for listening in the peak month, and not the six prior months. This decision limited the computation associated with estimating a model of this size and did not appear to affect model fit substantially according to tests we ran on subsamples of the data. The same data were used as in the previous analysis (untagged time series, of course, had values of zero for all possible tags). Formally, the regression model can be represented as follows, where again m is peak month, L is the number of listens in a given month, T_i is the binary indicator for a given tag, and f is the exponential-family function calculated by the GAM:

$$\log \sum_{i=1}^{6} L_{m+i} = b_0 + f(\log L_m) + \sum_{i=1}^{2{,}290} b_i T_i \tag{2}$$

After running the model, which explains approximately 28.5% of the variance in the data (adjusted R^2), 161 unique tags were statistically significant predictors at $\alpha = .001$, a threshold selected in order to account for the large number of comparisons against the null hypothesis being made in the regression model. We proceeded to examine which of these tags were relatively strong predictors in the model.

Unsurprisingly, most of the 161 tags tend to have a positive (albeit small) impact on future listening, as evidenced by positive regression coefficients and consistent with the small positive effect of tagging overall as found in the previous analysis. The most telling observation is that commonly-used genre tags (e.g. "pop", "jazz", and "hip-hop") tend to be weak positive predictors of future listening. In contrast, relatively strong predictors (both positive and negative) appear to be comparatively obscure, possibly idiosyncratic tags (e.g. "cd collection", "mymusic", "purchased 09").[5] To examine this trend quantitatively, we plot in Figure 2a the global tag usage (i.e. the total number of uses of a tag in our full dataset of approximately 50 million annotations) as a function of the tag's impact on listening indicated by its coefficient in the regression model. Similarly, we plot in Figure 2b the unique number of users utilizing the tag, again as a function of its regression coefficient. The value e^c, where c is a tag's regression coefficient, represents the number of listens we expect a user's post-listening behavior to increase or decrease by if she were to apply a (thus the strongest predictors lead to an increase of fewer than 7 listens on average). Finally, in each

[4] We chose a threshold of five to ensure that data was not too sparse for the regression model but was still inclusive of infrequently occurring tags. Again, qualitative results hold when using both more and less restrictive thresholds.

[5] For a full listing of the regression coefficients across all tags in the model, see https://dl.dropboxusercontent.com/u/625604/papers/lorince.joseph.todd.2015.sbp.supplemental/regression_coefficients.txt

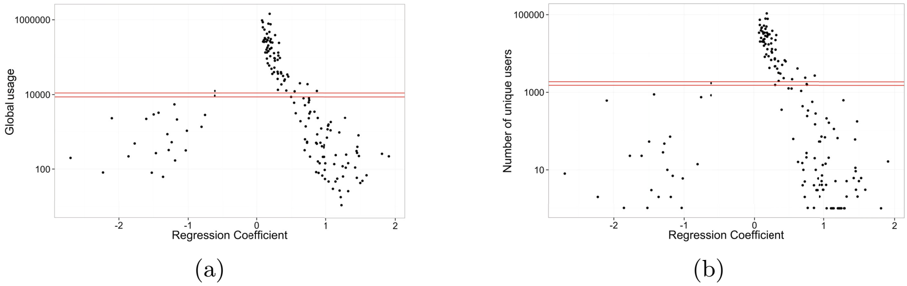


Fig. 2. In (a), each tag's global usage as a function of its regression coefficient; in (b) the number of unique users of each tag as a function of its regression coefficient

plot, the horizontal red bands mark the upper and lower limits of a bootstrapped 95% confidence interval on the popularity of the 2,129 remaining tags that were *not* significant in the regression model.

The data suggest that the most popular tags on both metrics (i.e., those above the red lines) are significant, weakly positive predictors of future listening, while relatively unpopular tags (i.e., below the red lines) tend to have relatively strong positive (and, in some cases, strong negative) impacts on listening. Tags which were not significant in the model (i.e., those that would fall between the red lines in each plot) had moderate popularity levels with respect to both metrics. The high statistical reliability but small regression coefficients for the most popular tags may be somwhat artifactual, primarily reflecting high variability in how predictive these tags are of listening across individuals. We believe this finding is still informative, however, as it indicates that popular tags are not consistently associated with future listening.

5 Conclusion

In this paper we set out to test the oft-cited assumption that tags serve as retrieval aids for individuals using collaborative tagging systems. We did so via a novel methodology, testing for evidence that tagging an artist increases a user's future listening to that artist in comparison to a carefully selected set of untagged time-series. Results suggest that tagging an artist does lead to an increase in listening, but that this increase is, on average, quite small (amounting to only 1 or 2 additional listens over a 6 month period). Given the various possible motivations for tagging, we expected only some tags to serve as retrieval cues, and thus tested the relative predictiveness of future listening for different tags. This analysis revealed systematic differences in how predictive the presence or absence of different tags was for future listening as a function of tag popularity. The data suggest that, at least for the small number of most highly significant tags that we consider, those that are globally popular have relatively little effect on future listening, and are generally associated with very small increases in post-taggging listening rates. The tags that seem to "matter" (i.e. those that are relatively strong predictors of whether or not a user will listen to an artist after tagging it) are generally much less popular. Even these stronger predictors, however, lead to relatively slight increases in listening. The strongest predictors are associated with a change of only about 7 listens over a six month period, on average.

Because we only analyzed a small sample of statistically influential tags, we are at this point tentative to make strong claims about which specific factors contribute to tags being better or worse predictors of increased listening, or even of decreased listening.[6] The evidence nevertheless suggests that relatively uncommon (and in many cases idiosyncratic) tags are most predictive of future listening behavior. The intriguing flipside is that the descriptive, popular tags that are arguably most useful to the community at large (i.e. genre labels and related tags) are not particularly associated with increases in listening for those who applied the tags, and thus are likely not functioning as memory cues.

Overall it appears that, while on average tagging an artist has a small positive effect on one's own future listening, the most common tagging activities are *not* strong predictors of future retrieval. We cannot be sure of the extent to which the many other possible tagging motivations are at play here, nor can we tell at this point if and when a tag is applied with the intention of being used for retrieval, while ultimately not being used for this purpose. That said, our results may indicate that the primary motivation for tagging on Last.fm is not for personal information management (tagging a resource for one's own retrieval), but rather may be socially or otherwise oriented, which may in turn result in tags that are useful for the community at large. This leads to the interesting possibility that a folksonomy can generate the useful, crowdsourced classification of content that proponents of collaborative tagging extol, even if this process is not strongly driven by the self-directed, retrieval-oriented tagging that is typically assumed in such systems.

While our results provide clues as to whether tags really function as retrieval aids, this remains early work addressing a hitherto unstudied research question. There is certainly room to refine and build upon the methods we present here for testing if and when tagging increases listening rates. In particular, our analysis at the level of artists (rather than the individual resources tagged) may be problematic, and we hope to develop models that operate directly at the level of the content tagged. There are also many factors we have not controlled for here that could be incorporated into future models, such as exogenous influences on listening (e.g. when an artist goes on tour or releases a new album), and we should explore alternative methods for normalizing and controlling for user listening habits beyond our approach here of simply considering raw monthly listens.

It will be critical to expand on methods for understanding which tags serve as memory cues and under what circumstances. It is clearly the case that not *all* tags function as memory cues, so more robustly identifying which tags do serve as retrieval aids is a fruitful direction for future work. Incorporating research on

[6] There were few enough of the latter in the current analysis that they may be largely due to statistical noise.

human memory from the cognitive sciences can also further inform hypotheses and analytic approaches to these questions, something we are actively pursuing in ongoing research. A final limitation is that we are exploring tagging in a particular collaborative tagging system, which operates in the possibly idiosyncratic domain of music. Tagging habits may vary systematically in different content domains, but until usable data becomes available, we can only speculate as to exactly how.

In closing, to address the titular question of whether or not tags function as retrieval aids, the best answer for Last.fm at least would appear to be "sometimes, but usually not". While there is much work to be done on when and why particular tags serve this function and others do not, it is clear that the overarching retrieval assumption is far from universally valid: Tags certainly do not always function as memory cues, and our results suggest that facilitating later retrieval may actually be an uncommon tagging motivation.

References

1. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 971–980. ACM (2007)
2. Cattuto, C., Loreto, V., Pietronero, L.: Semiotic dynamics and collaborative tagging. Proceedings of the National Academy of Sciences **104**(5), 1461–1464 (2007)
3. Glushko, R.J., Maglio, P.P., Matlock, T., Barsalou, L.W.: Categorization in the wild. Trends in Cognitive Sciences **12**(4), 129–135 (2008)
4. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science **32**(2), 198–208 (2006)
5. Gupta, M., Li, R., Yin, Z., Han, J.: Survey on social tagging techniques. ACM SIGKDD Explorations Newsletter **12**(1), 58–72 (2010)
6. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: Proceedings of the 16th International Conference on World Wide Web, pp. 211–220. ACM (2007)
7. Hastie, T.J., Tibshirani, R.J.: Generalized additive models (1990)
8. Heckner, M., Heilemann, M., Wolff, C.: Personal information management vs. resource sharing: towards a model of information behavior in social tagging systems. In: ICWSM (2009)
9. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: search and ranking. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006)
10. Körner, C., Kern, R., Grahsl, H.P., Strohmaier, M.: Of categorizers and describers: an evaluation of quantitative measures for tagging motivation. In: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, pp. 157–166. ACM (2010)
11. Lorince, J., Todd, P.M.: Can simple social copying heuristics explain tag popularity in a collaborative tagging system? In: Proceedings of the 5th Annual ACM Web Science Conference, pp. 215–224. ACM (2013)
12. Lorince, J., Zorowitz, S., Murdock, J., Todd, P.M.: "Supertagger" behavior in building folksonomies. In: Proceedings of the 6th Annual ACM Web Science Conference, pp. 129–138. ACM (2014)

13. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT 2006, tagging paper, taxonomy, flickr, academic article, to read. In: Proceedings of the Seventeenth Conference on Hypertext and Hypermedia, pp. 31–40. ACM (2006)
14. Vander Wal, T.: Explaining and Showing Broad and Narrow Folksonomies (2005). http://www.vanderwal.net/random/entrysel.php?blog=1635
15. Vander Wal, T.: Folksonomy Coinage and Definition (2007). http://vanderwal.net/folksonomy.html
16. Wood, S.N.: mgcv: GAMs and generalized ridge regression for R. R News **1**(2), 20–25 (2001)
17. Yeung, C.M.A., Noll, M.G., Gibbins, N., Meinel, C., Shadbolt, N.: SPEAR: Spamming-Resistant Expertise Analysis and Ranking in Collaborative Tagging Systems. Computational Intelligence **27**(3), 458–488 (2011)
18. Zollers, A.: Emerging motivations for tagging: expression, performance, and activism. In: Workshop on Tagging and Metadata for Social Information Organization, held at the 16th International World Wide Web Conference (2007)