

# Do tags really functions as retrieval aids?

Jared Lorince<sup>1</sup>  
jlorince@indiana.edu

Kenneth Joseph<sup>2</sup>  
kjoseph@cs.cmu.edu

Peter M. Todd<sup>1</sup>  
pmtodd@indiana.edu

<sup>1</sup>Cognitive Science Program  
Indiana University, Bloomington, Indiana, USA  
<sup>2</sup>Computation, Organization, and Society Program  
Carnegie Mellon University, Pittsburgh, PA, USA

## 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. Here we use a large-scale dataset from Last.fm to explore how patterns of music tagging and subsequent listening interact in an effort to determine if there exist measurable signals of tags functioning as retrieval aids. Results suggest that there exists only a small effect of tags increasing listening levels, with interesting differences in which kinds of tags are most associated with future listening. On the whole, our findings suggest that tagging generally is *not* associated with future retrieval.

## 1. INTRODUCTION

In social tagging systems, users assign freeform textual labels to digital content (music, photos, web bookmarks, etc.). It is widely assumed<sup>1</sup> in the literature that tags serve as personal 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 intrinsic to Vander Wal’s original definition of a folksonomy, which he contends “is the result of personal free tagging of information and objects...*for one’s own retrieval*” [3, emphasis added]. This perspective is echoed in many studies of tagging patterns. Is this assumption valid?

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 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

<sup>1</sup>Other researchers have proposed alternative motivations, however. Examples include personal information management, resource sharing, opinion expression, performance, and activism. See [1] for a review.

For the full version of this paper, see  
<https://dl.dropboxusercontent.com/u/625604/papers/Lorince.Joseph.Todd.SBP2015.pdf>

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: while 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 (in our case, music listening) with tagged versus untagged content vary. Our principal research question is whether listening patterns for tagged content are consistent with the expectation that tags serve as memory cues. If this were to be the case, 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.

Here we test this hypothesis using a large-scale dataset [2] consisting of the complete listening and tagging histories of 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 ask (a) does tagging an artist increases the probability of listening to that artist in the future, as shown by comparison of tagged versus untagged time-series? And (b) are certain tags particularly associated with increases in future listening?

## 2. ANALYSES AND RESULTS

*Comparison of tagged and untagged time-series* To provide a “fair” basis for comparing tagged and untagged data, we considered only tagged time series where a tag was applied in a user’s month of peak listening to an artist, temporally aligning each to that peak month. Untagged time series were aligned in the same manner. See the full paper for additional criteria used in sampling time series for analysis

We then developed a a Generalized Additive Model (GAM) model that incorporates previous listening behavior on post-peak behavior. Our dependent variable in the regression was the logarithm of the sum of all listens in the six months after a tag has been applied, to capture the possible effect of tagging over a wide temporal window. Note, however,

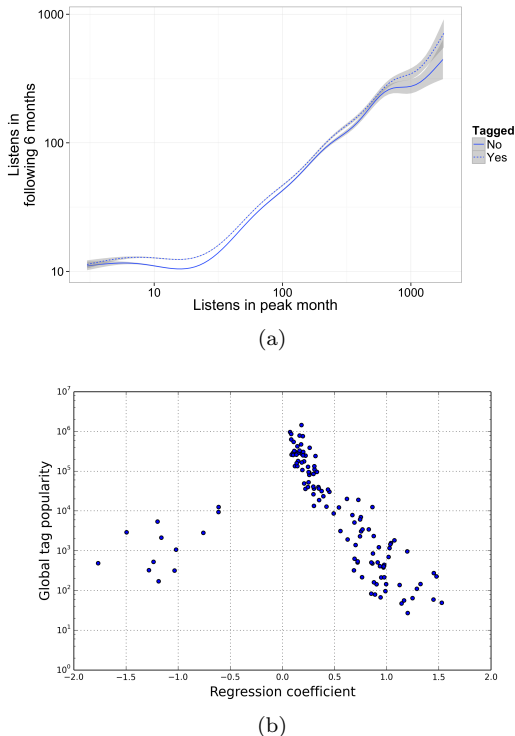


Figure 1: (a): Results of regression model 1. Shown is the predicted sum of listening in the 6 months after the peak (with 95% CI), as a function of the number of listens in the peak month. (b) Global popularity versus regression coefficient of tags in regression model 2, for the top 10% most reliable tags in the model (in all cases  $p < 0.001$ ).

that qualitative results hold when testing listening for each individual month as well. 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 the sum of listens in the peak month and the six previous months.

The regression model, which explained approximately 30% of the variance in the data (adjusted  $R^2$ ), indicated that the tagged/untagged indicator, as well as 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 1a instead displays a similar model which considers only the effect of listening in the peak month on post-peak listening. As this plot suggests and the full model confirms, we can conclude that, controlling for all previous listening behavior, a tag increases the logarithm of post-peak listens by .147 (95% CI = [.144, .150]). This indicates that 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.

**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. Instead of a single tagged/untagged indicator, however, we included binary (present / not present) regressors for all unique tags that had at least five occurrences in our subsample

The new model explained approximately 29% of the variance in the data (adjusted  $R^2$ ). To address our research question, we examined which particular tags were the strongest

predictors in the model. The most telling observation was that commonly-used genre tags (e.g. “pop”, “jazz”, and “hip-hop”) – which are the most common tags overall in our full dataset – tend to be positive, but very weak predictors of future listening. In contrast, relatively strong predictors appear to be comparatively obscure, possibly idiosyncratic tags (e.g. “cd collection”, “mymusic”, “purchased 09”). To examine this trend quantitatively, we plot in Figure 1b global tag popularity as a function of a tag’s impact on listening as indicated by its coefficient in the regression model.

The data suggest that the most popular tags under both metrics considered are weak predictors of future listening, while relatively unpopular tags tend to have either strong positive or strong negative impacts on listening.

### 3. CONCLUSION

This paper presents a novel methodology for 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 listens over a 6 month period). Given the various possible motivations for tagging, however, we expect 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 tags about which we can make statistically meaningful claims, those that are globally popular and well-known have relatively little effect on future listening. 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.

This suggests that, while on average tagging an artist has a small positive effect on future listening, the most common tagging activities are *not* strong predictors of future retrieval. We cannot be sure which of 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 oriented, resulting 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.

### 4. REFERENCES

- [1] M. Gupta, R. Li, Z. Yin, and J. Han. Survey on social tagging techniques. *ACM SIGKDD Explorations Newsletter*, 12(1):58–72, 2010.
- [2] J. Lorince, S. Zorowitz, J. Murdock, and P. M. Todd. “Supertagger” behavior in building folksonomies. In *Proceedings of the 6th Annual ACM Web Science Conference*, pages 129–138. ACM, 2014.
- [3] T. Vander Wal. Folksonomy Coinage and Definition, 2007.