Understanding the Motivation Behind Tagging

Christian Körner
Knowledge Management Institute
Graz University of Technology
Email: christian.koerner@tugraz.at

Abstract

Tagging is the process of annotating resources with keywords or terms in order to enable better organization, retrieval and sharing of the resource. Today, a large number of diverse applications exist that allow users to tag their documents. Del.icio.us, for example, enables users to organize their bookmarks and tag them accordingly. Flickr allows users to tag pictures and YouTube enables content creators to assign tags to videos in order to facilitate retrieval. While these systems have increasingly become an object of research for web scientists, little is known about these systems, for example how and why they are used, how behavior on a micro level propagates to phenomenon on the macro level, or we can simulate the dynamics of tagging system's evolution. My research interest lies in understanding and modeling the motivation of users, in particular why and how they are using these systems. To this purpose I introduce measures which support the examination of a users tagging history and enable the classification of users into two vastly different types of tagging motivation: *categorizers* and *describers*. In an evaluation I show how tagging motivation differs on various social tagging platforms and what influences the motivation types have on folksonomies.

Introduction and Related Work

Tagging typically describes the voluntary activity of users who are annotating resources with terms (so called "tags") freely chosen from a freely chosen and uncontrolled vocabulary (cf.[Golder2006],[Marlow2006]). In social tagging systems, the structure of data captured by tagging can be characterized by a tripartite graph with hyper edges. The three disjoint, finite sets of such a graph correspond to 1) a set of persons or users $u \in U$ 2) a set of objects or resources $r \in R$ and 3) a set of annotations or tags $t \in T$ that are used by users U to annotate resources R. A folksonomy is defined as the set of annotations $F \subseteq U \times T \times R$ (cf. [Mika2007]). A personomy F_u of a user $u \in U$ is the reduction of F to u [Hotho2006].

Social tagging systems have been the subject of research for several years now. The influential paper by [Mika2007] for example describes how social tagging systems can be used for ontology learning. Other work has studied the evolution of folksonomies over time and using tags for a variety of other purposes (such as enhancing web search etc.).

However, our understanding of folksonomies is not yet (nor could it be) mature. A question that has recently attracted the interest of our community is whether the properties of tags in tagging systems and their usefulness for different purposes can be assumed to be *a function of the taggers' motivation or intention behind tagging* [Heckner2009]. If this was the case, tagging motivation would have broad implications. In order to assess the general usefulness of algorithms that aim to - for example - capture knowledge from folksonomies, we would need to know whether these algorithms produce similar results across user populations driven by different motivations for tagging. Recent research already suggests that different tagging systems afford different motivations for tagging [Heckner2009]. Further work presents anecdotal evidence that even within the same tagging system, motivation for tagging between individual users may vary greatly [Wash2007]. Given these observations, it is interesting to study whether and how the analysis of user motivation for tagging is amenable to quantitative investigations, and whether folksonomies and the tags they contain are influenced by different tagging motivations.

Tagging motivation has remained largely elusive until the first studies on this subject have been conducted in 2006. At this time, the work by [Golder2006] and [Marlow2006] have made advances towards expanding our theoretical understanding of tagging motivation by identifying and classifying user motivation in tagging systems. Their work was followed by studies proposing generalizations, refinements and extensions to previous classifications [Heckner2009]. An influential observation was made by [Coates2005] and elaborated on and interpreted in [Marlow2006] and [Heckner2009]. This line of work suggests that a distinction between at least two fundamental types of user motivation for tagging is important:

On one hand, users who are motivated by categorization view tagging as a means to *categorize resources* according to some shared high-level characteristics. These users tag because they want to construct and maintain a navigational aid to the resources for later browsing. On the other hand, users who are motivated by description view tagging as a means to accurately and precisely *describe resources*. These users tag because they want to produce annotations that are useful for later searching. This distinction has been found to be important because, for example, tags assigned by describers might be more useful for information retrieval (because these tags focus on the content of resources) as opposed to tags assigned by categorizers, which might be more useful to capture a rich variety of possible interpretations of a resource (because they focus on user-specific views on resources).

Two Different Types of Tagging Motivation

In this work we differentiate two vastly different types of tagging motivation - *categorizers* and *describers* - which are depicted in greater detail in the following sections:

Categorizers - Using Tags to Categorize Resources

Users who are motivated by categorization engage in tagging because they want to construct and maintain a navigational aid to the resources (URLs, photos, etc) being tagged. This implies developing a limited set of tags (or categories) that is rather stable over time. The tags that are assigned to resources are regarded as an investment into a structure, and changing this structure is typically regarded to be costly to a categorizer. Resources are assigned to tags whenever they share some common characteristic important to the mental model of the user (e.g. 'photos', 'trip to Vienna' or 'favorites'). Because the tags assigned by categorizers are very close to the personal preferences and ideas of users, they can act as suitable facilitators for navigation and browsing.

Figure 1: Example Tag Cloud of a Categorizer

Describers - Using Tags to Describe Resources

Users who are motivated by description engage in tagging because they want to accurately and precisely describe the resources being tagged. This typically implies an open set of tags, with a rather dynamic and unlimited tag vocabulary. Tags are not viewed as an investment into a tag structure, and changing the structure continuously is not regarded as costly. The goal of tagging is to identify those tags that match the resource best. Because the tags assigned are very close to the content of the resources, they can be utilized for searching and retrieval.

Figure 2: Example Tag Cloud of a Describer

Figures 1 and 2 show examples of typical tag clouds of categorizers and describers. Table 1 gives an overview of main intuitions about the two types of tagging motivation.

Table 1: Intuitions about Categorizers and Describers

	Categorizer	**Describer**
Goal	Later Browsing	Later Retrieval
Change of Vocabulary	costly	cheap
Size of Vocabulary	limited	open
Tags	subjective	objective
Tag Reuse	frequent	rare
Tag Purpose	mimicking taxonomy	descriptive labels

Originality of Approach

In the related work it can already be seen that understanding the motivation a user of a software system has is nothing new. However previous work has tried to get information on the intentions users of tagging systems by empirical studies such as questionnaires, human subject studies, interviews or expert judgements. To the best of our knowledge this is the first approach which tries to automatically evaluate a user's behavior by examining the tagging history.

Data Sets

For our experiments we crawled a number of diverse tagging datasets which had the following requirements:

- Each of the captured personomies had to contain at least 500 tagged resources (posts).
- All captured personomies had to be complete instead of capturing only fractions.
- The captured posts had to be in chronological order to enable the detection of potential shifts in tagging motivation.

The following datasets were investigated in our experiments:

- Flickr - An online photo sharing system
- Del.icio.us - A social bookmarking system
- Bibsonomy - A Social Publication and Bookmark Sharing System
- CiteULike - A Social Publication Sharing System
- Diigo - A social bookmarking system which also supports the creation of lists.
- MovieLens - An online movie recommendation system which also supports tagging.

The crawled datasets contain two "synthetic" datasets as well as eight "real world" datasets. The "synthetic" datasets were chosen in order to simulate "extreme" taggers according to their tagging motivation. The first "synthetic" dataset is the ESP game which was used to establish the notion of a "perfect describers" because players of this game need to find the most descriptive keywords for pictures. The second synthetic dataset captures the fact that a picture is typically only stored in one photo set and therefore mimics personomies of "perfect categorizers".

Table 2 gives an overview of the datasets acquired for this research. To the best of our knowledge, the aggregate collection of these datasets itself represents the most diverse and largest dataset of complete and very large personomies (>500 tagged resources, recorded from the user's first bookmark on) to date.

Table 2: Properties of the Crawled Data Sets

Dataset	|U|	|T|	|R|	$\|R\_u\|_{min}$	|T|/|R|
ESP Game*	290	29,834	99,942	1,000	0.2985
Flickr Sets*	1,419	49,298	1,966,269	500	0.0250
Delicious	896	184,746	1,089,653	1,000	0.1695
Flickr Tags	456	216,936	965,419	1,000	0.2247
Bibsonomy Bookmarks	84	29,176	93,309	500	0.3127
Bibsonomy Publications	26	11006	23696	500	0.4645
CiteULike	581	148,396	545,535	500	0.2720
Diigo Tags	135	68,428	161,475	500	0.4238
Movielens	99	9,983	7,078	500	1.4104

Measures for Detection of Tagging Motivation

Several measures were developed to detect a user's motivation for tagging. These measures should capture the usage of tags instead of evaluating their semantics.

Vocabulary Size

Intuition: Over time, the tag vocabulary (i.e. the set of all tags used in a users' personomy) of an ideal categorizer's tag vocabulary would be expected to reach a plateau. The rationale for this idea is that only a limited set of categories is of interest to a user mainly motivated by categorization. On the other hand, an ideal describer would not aim to limit the size of her tagging vocabulary, but would rather actively introduce new tags as required by the contents of the resources being tagged.
Description: A measure to distinguish between these two approaches would examine whether the cumulative tag vocabulary of a given user plateaus over time.

$$M_{vocab} = \frac{|T|}{|R|} \qquad (1)$$

Orphan Ratio

Intuition: Categorizers would have a low interest in introducing orphaned tags, i.e. tags (or categories) that are only used for a single resource, and are then abandoned. Introducing categories of size one seems counterintuitive for categorizers, as it would prevent them from using their set of tags for browsing efficiently. On the other hand, describers would not refrain from introducing

orphaned tags but would rather actively introduce a rich variety of different tags to resources so that re-finding becomes easier when searching for them later. Following this intuition, describers can be assumed to have a tendency to produce more orphaned tags than categorizers.
Description: The number of tags that are used only once within each personomy in relation to the total number of tags would be a simple measure to detect this difference in motivation. A more robust approach to measuring the influence of orphaned tags is described as follows:

$$M_{desc} = \frac{|\{t : |R(t)| \leq n\}|}{|T|}, n = \lceil \frac{|R(t_{max})|}{100} \rceil \qquad (2)$$

Conditional Tag Entropy

Intuition: For categorizers, useful tags should be maximally discriminative with regard to the resources they are assigned to. This would allow categorizers to effectively use tags for navigation and browsing. This observation can be exploited to develop a measure for tagging motivation when viewing tagging as an encoding process, where entropy can be considered a measure of the suitability of tags for this task. A categorizer would have a strong incentive to maintain high tag entropy (or information value) in his tag cloud. In other words, a categorizer would want the tag-frequency distribution as equally distributed as possible in order for her to be useful as a navigational aid. Otherwise, tags would be of little use in browsing. A describer on the other hand would have little interest in maintaining high tag entropy as tags are not used for navigation at all.
Description: In order to measure the suitability of tags to navigate resources, we develop an entropy-based measure for tagging motivation, using the set of tags and the set of resources as random variables to calculate conditional entropy. If a user employs tags to encode resources, the conditional entropy should reflect the effectiveness of this encoding process:

$$H(R|T) = -\sum_{r \in R} \sum_{t \in T} p(r,t) log_2(p(r|t)) \qquad (3)$$

$$M_{cat} = \frac{H(R|T) - H_{opt}(R|T)}{H_{opt}(R|T)} \qquad (4)$$

Combined Measure

The final measure is a combination of the two previous measures using the assumption about a describers typical behavior and the categorizers way to organize tags. This measure is the arithmetic mean of the two measures (see formula 5).

$$M_{combined} = \frac{M_{desc} + M_{cat}}{2} \qquad (5)$$

Figure 3 shows the four introduced measures applied onto three of our data sets and shows that users found in normal social tagging systems are found between the two reference data sets: The ESP game for simulating extreme describer behavior and the Flickr Sets Dataset for mimicking categorizer behavior.

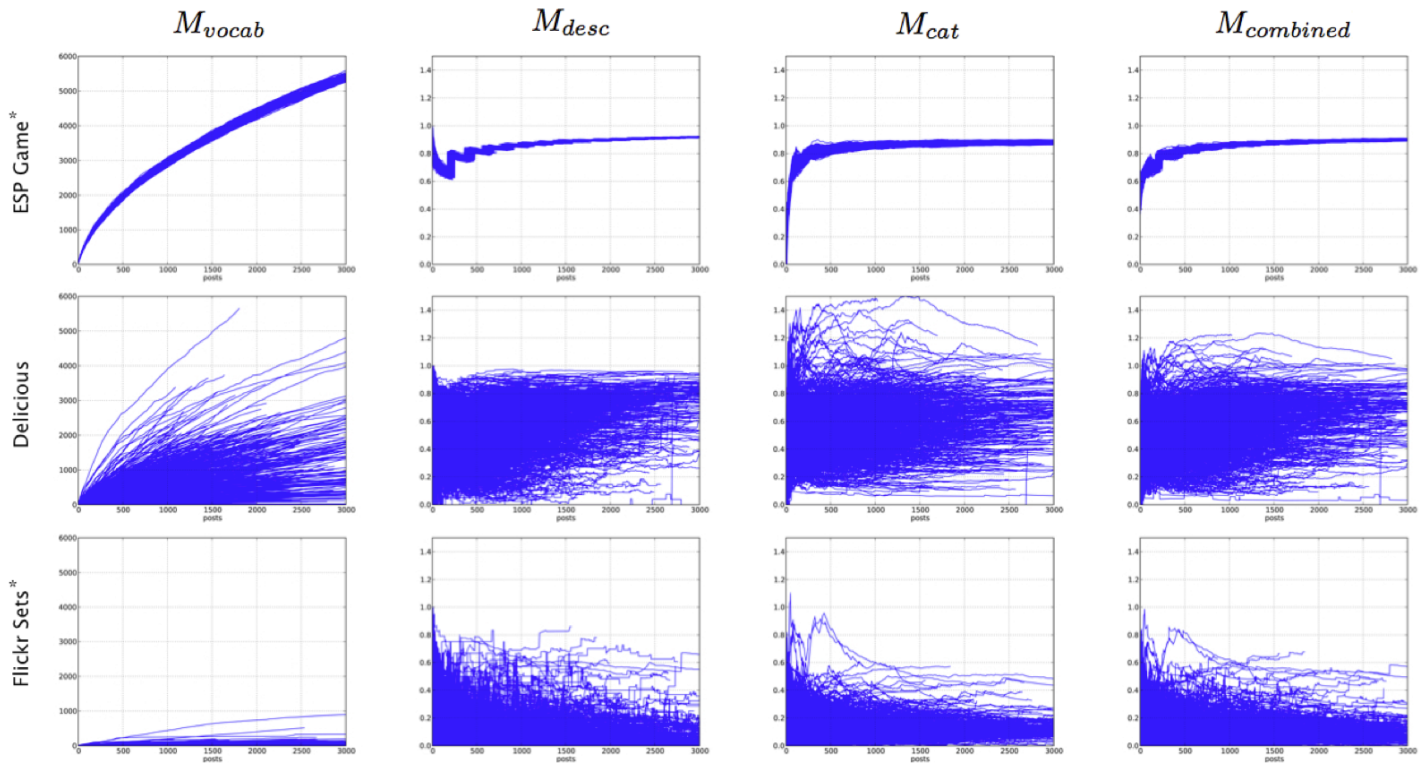

Figure 3: Overview of the introduced measures (from left to right: M_{vocab}, M_{desc}, M_{cat} and $M_{combined}$) over time for the two synthetic datasets (top and bottom row) and the Delicious dataset (middle row)

Results

In this section selected results of my research are presented.

Tagging Motivation Varies Across Tagging Systems

An initial experiment evaluated how the different networks differ in their users' tagging motivation. Figure 4 shows the seven datasets and their difference in tagging behavior. The calculations were done with the combined measure. The two rows in the back represent the two synthetic extreme datasets. ESP game as a reference for typical describer behavior and Flickr Photo Sets as a reference for typical categorizer behavior. An interesting finding is that tagging motivation varies across different social tagging platforms.

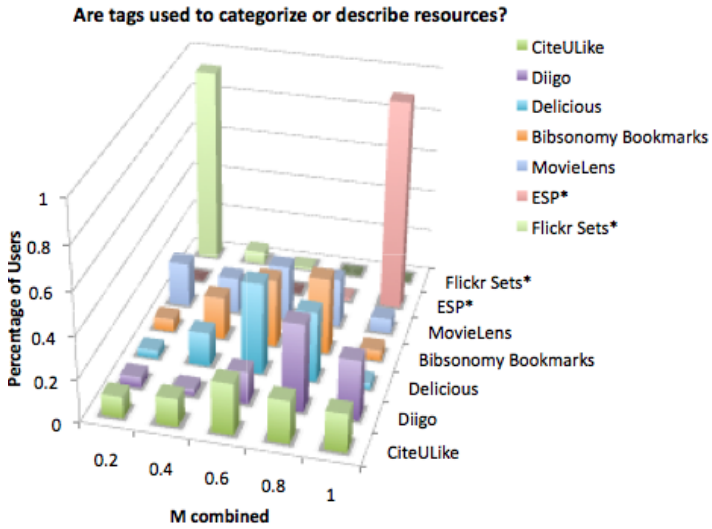

Figure 4: M combined at |Ru|=500 for 7 different datasets, binned in the interval [0.0 .. 1.0].

The two back rows reflect opposite extreme behaviors.

Differences in Tag Agreement

In an additional experiment, we examined whether the intuition that describers agree on more tags is correct. For this purpose we divided the users of our Del.icio.us data set in groups of equal size. Users who had a $M_{combined}$ value lower than 0.5514 were referred to as *Delicious Categorizers* whereas users with a higher value were denoted *Delicious Describers*. For each of the two groups we generated a tag set of the 500 most popular resources. For both of these tag sets we calculated the tag agreement, i.e. the number of tags that *k* percent of users agree on for a given resource.

Table 3: Tag agreement among Delicious describers and categorizers for 500 most popular resources. For all different *k*, describers produce more agreed tags than categorizers.

k	10	20	30	40	50	60	70	80
Desc. Wins	379	464	471	452	380	287	173	69
Cat. Wins	56	11	5	7	5	3	4	4
Ties	65	25	24	41	115	210	323	427

Table 3 shows the agreement values of *k* percent of users. We restricted our analysis to $T_u > 3$ in order to avoid irrelevant high values in this calculation. In all cases - for different values of *k* - describers produce more agreed tags than categorizers.

Human Subject Study

We evaluated whether individual users that were identified as extreme categorizers / extreme describers by $M_{combined}$ were also confirmed as such by human subjects. In our evaluation, we asked one human subject (who was not related to this research) to classify 40 exemplary tag clouds into two equally-sized piles: a categorizer and a de- scriber pile. The 40 tag clouds were obtained from users in the Delicious dataset, where we selected the top20 categorizers and the top20 describers as identified by $M_{combined}$. The inter-rater agreement kappa between the results of the human subject evaluation and $M_{combined}$ was 1.0. This means the human subject agrees that the top20 describers and the top20 categorizers (as identified by $M_{combined}$) are good examples of extreme categorization / description behavior.

Contributions

Our research highlights several opportunities for designers of social tagging systems to influence user behavior. While categorizers could benefit from tag recommenders that recommend tags based on their individual tag vocabulary, describers could benefit from tags that best capture the content of the resources. First experiments on the topic of tag recommendation were done in [Koerner2010b]. Offering users tag clouds to aid the navigation of their resources might represent a way to increase the proportion of categorizers, while offering more sophisticated search interfaces and algorithms might encourage users to focus on describing resources.

Furthermore sub-folksonomies containing taggers motivated by either motivation seem to be better for tasks like knowledge acquisition. This is an observation was exploited by [Koerner2010a]. In this work we showed that "verbose" taggers (in this case *describers*) are most useful of the emergence of tag semantics. Only a fraction (about 40%) of users of a social tagging system is needed to match or even outperform the semantic precision obtained from the complete dataset.

Conclusion and Future Work

This work introduced a quantitative way for measuring and detecting the tacit nature of tagging motivation in social tagging systems. We have evaluated these measures with synthetic datasets of extreme behavior as points of reference, via a human subject study and via triangulation with previous findings. Based on a large sample of users, our results show that 1) tagging motivation of individuals varies within and across tagging systems, and 2) that users' motivation for tagging has an influence on resulting tags and folksonomies. By analyzing the tag sets produced by Delicious describers and Delicious categorizers, we showed that agreement on tags among categorizers is significantly lower compared to agreement among describers.

In the future I plan to study tag recommendation strategies with regards to the current two tagging motivation types. Additionally I want to investigate other types of tagging motivation.

Acknowledgements

The research presented in this work is in part funded by the Know-Center and the FWF Austrian Science Fund Grant P20269. The Know-Center is funded within the Austrian COMET Program under the auspices of the Austrian Ministry of VIT, Austrian Ministry of WA and by the State of Styria.

References

[Coates2005] Coates, T. 2005. Two cultures of fauxonomies collide, http://www.plasticbag.org/archives/2005/06/two_cultures_of_fauxonomies_collide/. Last access: May 8:2008.

[Dellschaft2008] Klaas Dellschaft and Steffen Staab, An epistemic dynamic model for tagging systems, in 'HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia', ACM, New York, NY, USA, pp. 71--80, (2008).

[Golder2006] Golder S. and Huberman B.; Usage Patterns of Collaborative Tagging Systems; Journal of Information Science; 32(2):198, 2006

[Heckner2009] M. Heckner and M. Heilemann and C. Wolff. Personal Information Management vs. Resource Sharing: Towards a Model of Information Behaviour in Social Tagging Systems. Int'l AAAI Conference on Weblogs and Social Media ICWSM, San Jose, CA, USA, 2009.

[Hotho2006] Andreas Hotho and Robert Jäschke and C. Schmitz and Gerd Stumme, BibSonomy: A social bookmark and publication sharing system. In Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th Int'l Conference on Conceptual Structures, 87–102. 2006.

[Körner2010a] Christian Körner and Dominik Benz and Markus Strohmaier and Andreas Hotho and Gerd Stumme. Stop Thinking, start Tagging - Tag Semantics emerge from Collaborative Verbosity. Proceedings of the 19th International World Wide Web Conference WWW 2010, ACM, Raleigh, NC, USA, 2010. (to appear)

[Körner2010b] Christian Körner and Roman Kern and Hans Peter Grahsl and Markus Strohmaier. Of Categorizers and Describers: An Evaluation of Quantitative Measures for Tagging Motivation. 21st ACM SIGWEB Conference on Hypertext and Hypermedia HT 2010, ACM, Toronto, Canada, 2010. (to appear)

[Marlow2006] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia, pages 31–40, New York, NY, USA, 2006. ACM.

[Mika2007] Peter Mika, Ontologies are us: A unified model of social networks and semantics. Web Semantics: Science, Services and Agents on the World Wide Web 5(1):5–15, 2007.

[Wash2007] Rick Wash and Emiliee Rader, 2007. Public bookmarks and private benefits: An analysis of incentives in social computing. In ASIS & T Annual Meeting. 2007.