Analysing User Motivation in an Art Folksonomy

Max Arends, Josef Froschauer, Doron Goldfarb, Dieter Merkl and Martin Weingartner
Institut für Softwaretechnik und Interaktive Systeme
Technische Universität Wien
Favoritenstrasse 9-11, 1040 Wien, Austria
vsem@ec.tuwien.ac.at

ABSTRACT

The perception of art is a subjective affair - being influenced by our feelings, education and cultural background. Contrary, the study of art history uses formal methods to classify artworks. This discrepancy often poses a risk of being insurmountable – especially for users without prior knowledge of art history. The concept of social tagging provides the possibility to merge art historical information with the subjective perception of users. For our art Web platform *explorARTorium*, social tags augment exiting art historical information. In order to better understand how social tagging is best applied, it is necessary to examine the user's motivation to assign tags. We adopt the differentiation between users who are motivated by categorizing, and users who are motivated by describing resources. By evaluating our folksonomy according to this paradigm, we show that the preference for certain artworks has an effect on the user's tagging motivation, whereas the presentation of an artwork does not. While measures exist that are able to identify the user's motivation for annotating artworks, we propose an heuristic that aims to classify categorizing, respectively descriptive, tags. After evaluating this proposed heuristic, we show that it is indeed possible to identify categorizing and descriptive tags, even though the results are somewhat biased by the content of the resources and the individual tagging behaviour of the users.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: *Human Factors, Human Information Processing*; H.3.7 [**Digital Libraries**]: *Collection, User Issues*; J.5 [**Arts and Humanities**]: *Fine Arts*

General Terms

Human Factors, Design

Keywords

social tagging, folksonomy-mining, user interaction, cultural heritage, user profiling

i-Know '12 , September 05 - 07 2012, Graz, Austria

1. INTRODUCTION

Social tagging is used for the annotation of content for different kinds of resources like photos, videos, bookmarks, scientific papers, persons, geographical locations – and artworks. This list may be extended to any content that can be augmented with user annotated keywords (tags). For the domain of art, social tagging provides several advantages. Firstly, the concepts of art, relations and influences among artworks and artists, as well as iconographic metaphors are difficult to access for people without sophisticated knowledge in art history. Browsing art collections by following overlapping tags and the representation of different epochs and styles in visualizations like tag-clouds, may aid users to develop a better understanding for art historical concepts. Secondly, the presented information about artworks often uses a systematic, artificial language of experts that excludes users who are unfamiliar with that specific language [19]. Social tagging provides information by the users for the users, that can improve browsing and searching and serves as an additional layer of information description [17] [7]. Thirdly, by closely examining and annotating artworks, users are able to express their knowledge and feelings, which thereby further fosters their interest in art [11].

Several platforms exist that allow users to browse through art collections and annotate artworks. The steve.museum[1] project, a tagging platform that was created in collaboration of several US/UK museums, encourages users to tag artworks of their collections [15] [16] [17]. Flickr Commons[2] is a Web platform where cultural institutions publish artworks, mostly photographs, under the creative commons license and allow the Flickr-community to tag, comment and share the published artworks [13]. With the art-tagging-game artigo[3], users assign tags for artworks while competing against other users [4]. We created our own art Web platform, the explorARTorium[4], that combines methods of visual exploratory search, contextual view, social tagging and social media [2] [3].

Sen et al. [12] collapsed seven classes of tags from Golder and Hubermann [5] into three general classes: *factual tags*, *subjective tags* and *personal tags*. *Factual tags* identify the "facts" of a resource. For the domain of artworks, examples are the subject, details, motives and locations. The thereby annotated tags are commonly agreed on and therefore objective. *Subjective tags* express user opinions and interpre-

[1]http://www.steve.museum
[2]http://www.flickr.com/commons
[3]http://www.artigo.org
[4]http://www.explorARTorium.info

tations of the artwork. They describe the subjective view or even the feelings that are evoked by perceiving an artwork. Examples for subjective tags are *gloomy*, *friendly* or *wonderful*. *Personal tags* describe the personal relation that users have with an artwork, for example *ilike* or *myfavorites*. The proportion of factual, subjective and informative tags varies greatly for different resources and personomies.
In order to better understand how social tagging works, it is necessary to understand the motivations why and how users annotate tags. Depending on the resource, the user's knowledge and the user-interface, the collected tags vary greatly. Körner et al. [7] [8] [9] show that the motivation behind tagging has an significant impact on how resources are tagged. By gaining a deeper understanding for the user motivation, user-interfaces for social tagging can be enhanced. Lots of different motivations for assigning tags exist. For the sake of the scope of this paper we will only distinguish between two main motivations introduced by Körner et al. [8] [9] and Strohmaier et al. [14]: categorization and description.

This paper examines existing heuristics to distinguish the user's motivation behind assigning tags for artworks and evaluates the art-folksonomy collected with the explorAR-Torium. Since there is not only a difference between the user's motivation, but also among the tags themselves, we introduce and evaluate a method that identifies categorizing and descriptive tags.

This paper is organized as follows: In Section 2 we give an overview of related work and describe ways to identify already existing heuristics to identify user motivation. In Section 3 we introduce an heuristic to identify categorizing and descriptive tags. In Section 4 we introduce our experimental setup. In Section 5 we show how our tagging-environments influenced the user's motivation and evaluate our proposed measure to distinct between categorizing and descriptive tags. Finally, Section 6 concludes the paper.

2. RELATED WORK

Marlow et al. [10] state that the tagging behaviour varies greatly according to the system design and the user's motivation for tagging a resource. Golder and Huberman [5] come to the same conclusion by analysing categories of tags that describe bookmarks. Hammond et al. [6] and Heckner et al. [7] describe the differences of the users motivations for different tagging environments. Based on these works, Körner et al. [8] distinguish between two mayor sets of users:

- categorizers: users who's motivation is to categorize resources. They assign more general terms that describe concepts. Categorizers have a smaller tagging-vocabulary and often reuse the same tags. The resulting tags can therefore be for used as some kind of classification.
- describers: users who provide a very detailed description of the resources. The resulting tags vary a lot, are rarely reused and contain a great amount of synonyms. Descriptive tags are very useful for later browsing and searching.

Körner et al. describe several measures in [8] [9] to identify categorizers and describers. Since some measures in [9] are found to provide equally good results, we present only heuristics that are best suited to distinguish the user's motivation for our tagging environment and our folksonomy.

To clarify the following heuristics, we define a $folksonomy$ according to [8] as a tuple $F := (U, T, R, Y)$ with a finite set of users (U), who assign tags (T) to resources (R). Y is a ternary relation between them, i.e., $Y \subseteq U \times T \times R$. Therefore each tag assignment is a triple (u, T_{ur}, r) of a user, a tag and a resource with $u \in U$ and $r \in R$. The complete set of tags of a user (T_u) is called personomy. R_u is the set of resources that are tagged by a user.

vocabulary size – $vocab(u)$

Since describers tend to annotate a resource in far more detail than categorizers, their vocabulary is much larger than the vocabulary of a categorizer, who often reuses the same tags. This measure does not take the number of annotated resources into account. Furthermore, it is difficult to compare personomies of different sizes.

$$vocab(u) = |T_u|$$

tag/resource ratio – $trr(u)$

The vocabulary size of a user is set into relation with the total number of annotated resources of a user. Users with less distinct tags for a resource will more likely be categorizers as they are aiming for a more general description of a resource. A describer would annotate more details and therefore use more distinct tags.

$$trr(u) = \frac{|T_u|}{|R_u|}$$

average tags per post – $tpp(u)$

This measure is an improvement of the tag/resource ratio, as it also takes into account how many tags a user assigns to a resource on average. Users that assign more tags to a resource are therefore more likely describers, since they tend to annotate the resource in much more detail.

$$tpp(u) = \frac{\sum^{r} |T_{ur}|}{|R_u|}$$

orphan ratio – $orphan(u)$

Contrary to categorizers who use a fixed vocabulary, describers tend to annotate resources in great detail. Therefore describers do not reuse tags as often as categorizers. The orphan ratio relates the amount of seldom assigned tags to the total amount of assigned tags of a user. t_{max} defines the value of the most assigned tag of a user. The orphan tags T_u0 define all tags that are seldom assigned to resources. These tags are then related to the total amount of tags of a user (T_u), in order to calculate the orphan ratio. Users that produce a high amount of orphaned tags are more likely describers, since they annotate more details. The orphan ratio ranges between zero and one. A score close to one identifies a user with a great amount of orphaned tags and therefore a describer. We adopted the orpan ratio in a way, that if the number of assigned tags n is zero, we assume n to be one, in order to get adequate results.

$$orphan(u) = \frac{|T_u0|}{|T_u|}, Tu0 = \{t||R(t)| \leq n\}, n = \lceil \frac{R(t_{max})}{100} \rceil$$

The presented measures to distinguish the tagging behaviour in categorizers and describers are solely based on the user behaviour and the pragmatics of annotating resources. They do not focus on identifying the tags themselves.

3. CLASSIFICATION OF TAGS

In an ideal world, there would be a strong distinction between categorizers and describers, and according to their motivation they would use tags solely to categorize or describe a resource. But in the real world (or at least in our tagging environment), things are not that simple. For example, a categorizer who assigns the tag *woman* to an artwork, might be motivated by identifying all artworks that portray women, whereas a describer might use the tag *woman* to describe a small detail of an artwork. By analysing the vocabulary from the users of our tagging environment, we see that most users represent a mixture of categorizers and describers. Categorizing tags are assigned for artworks that users have knowledge about, and depending on their perception of the artwork, they use descriptive tags to annotate further details. For the example of the explorARTorium, users tend to provide categorizing tags like *painting*, *statue*, *bust* or *renaissance*. Additionally, the same user might assign other resources with descriptive tags like *beard*, *blue eye*, etc. In-between is a great amount of tags that cannot be accordingly classified as categorizing or descriptive, since they represent both motivations. Therefore, we propose an heuristic that aims to distinguish between tags that are mostly used for categorization, and tags that are mostly used for description.

average amount of tags per post – $atpp(t)$

This measure is motivated by the definition of Körner et al. [9] that describers annotate a resource in great detail and verbosely, while categorizers aim to annotate a resource efficiently. Therefore, describers assign a lot more tags to a resource than categorizers.
The $atpp(t)$ heuristic calculates the average amount of tags that were assigned to a resource together with a specific tag. To be more precisely, the $atpp(t)$ measure of a specific tag is calculated as the amount of distinctive tags that were separately annotated by all users (that assigned the tag t) for all resources (that are annotated with the tag t) together with the specific tag, divided by the number of distinctive resources annotated with the specific tag t.

$$atpp(t) = \frac{\sum_{r \in R_t} \sum_{u \in U_t} |T_{ur}|}{\sum_{u \in U_t} |R_{ut}|}$$

A low $atpp(t)$ score therefore indicates that a tag is used for categorization, while tags used for description get high scores. Tags that get average scores are not classified, since they might be used for both, categorization and description. Since it is possible that a user might assign a categorizing tag to a resource, for example the tag *painting*, and then further annotates the resource with descriptive tags like *hat* or *beard*, a high number of resources is needed to estimate the motivation behind a tag for the whole folksonomy. Therefore, the $atpp(t)$ measure is best applied for tags that are frequently used. Also, as some users are rather categorizers than describers and vice-versa, the $atpp(t)$ heuristic should only be applied to tags that were assigned by multiple users. For our analysis, the $atpp(t)$ measure is applied to tags that are assigned to over ten different resources ($R_t > 10$) and that have been annotated by at least three different users ($U_t \geq 3$).

4. EXPERIMENTAL SETUP

We examine the folksonomy of the explorARTorium, a Web platform that combines methods of visual exploratory search, contextual view, social tagging and social media. For more information about the concepts of the explorARTorium, we refer to our previously published work in [2] [3]. In order to collect tags for the explorARTorium, we created the TaggingTool[5], which focused solely on collecting user generated content for artworks. In order to collect tags in an efficient and unbiased way, we created the TaggingTool as a Web site that randomly displays an artwork without any textual or visual description. Due to the absence of additional information, users had to assign tags based on their own perception and knowledge, as well on the feelings they had by reckoning the works of art. Additionally, the users were given the possibility to rate the artwork on a scale from zero to five. As a way to boost the user's motivation, the total amount of their assigned tags was constantly calculated, and presented as a high-score list together with the amount of tags of their predecessor and successor.
Figure 1 shows a screenshot of the explorARTorium and the TaggingTool.
For our data source, we extracted all open available information about the artworks, and downloaded the pictures, from the Web Gallery of Art (WGA), which is a Web database for mainly European art from the 11th to 19th century. By the time our snapshot of the WGA was taken, information about 12,742 paintings by 1,542 different artists from a timespan between 1100 to 1900 AD was available. The WGA offers basic meta-information about each artwork. Besides the information about the title, artist, region and date of creation for an artwork, the WGA relates each artwork to a certain theme. These themes are *religious*, *historical*, *portrait*, *landscape*, *genre*, *mythological*, *still-life*, *historical*, *interior*, *study* or – if it doesn't fit in any category – *other*.

4.1 Different Tagging Environments

Social tags were collected in three different environments:

V1 : October 2010 - November 2010
The first version of the TaggingTool did not offer the possibility to view tags of other users. Only the artwork itself was displayed, without further information or tags. We therefore hoped to get consensual tags for an artwork, in order to collect a narrow folksonomy [18] and identify tags that were assigned by multiple users. Since the users were not influenced by other tags or textual/visual description for the artwork, the annotated tags can be regarded as the original vocabulary of the users. "Original" therefore means that it represents solely the knowledge and perception of the users, without influence by the annotations of other users. In further analysis this vocabulary is compared to tags that were annotated while the users had the possibility to view tags of other users.

[5] http://vsem.ec.tuwien.ac.at/taggingtool

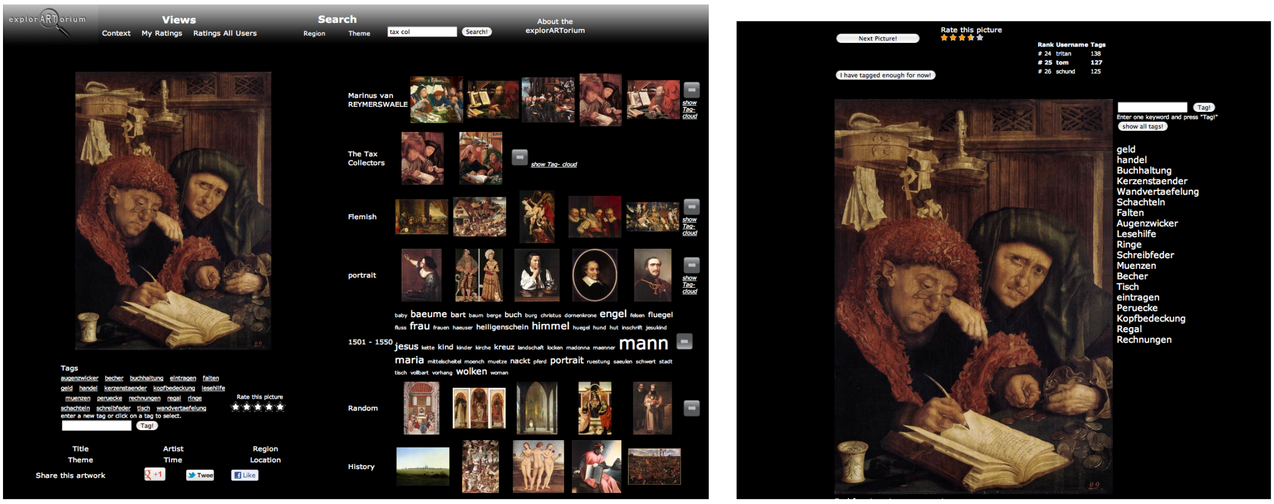

Figure 1: screenshots of the explorARTorium (left) and the Taggingtool (right).

V2 : December 2010 - March 2011
The changes in V2 aimed to get a more detailed description, which we hoped would provide additional entry points for the exploration of the art collection. The main difference to V1 was that already assigned tags could be viewed. We thereby intended to get users to examine artworks more closely, and thereby identify the content of the tags and then annotate further details. Instead of displaying the tags natively, the users had to actively demand to view them by pressing a button. By logging this action, it is possible to identify tags that were annotated additionally to already assigned tags. Also, it is possible to identify which tags the users actively perceived, since we assume that the users did indeed read the tags after demanding to see them. In further analysis this vocabulary is compared to tags that were annotated while additional information was presented for the artworks.

V3 : March 2011 - now
The explorARTorium uses a completely new user interface. Instead of displaying artworks in random order, the explorARTorium allows users to actively select artworks they want to explore and possibly tag. Additional textual information about the artworks are presented and further artworks are visually displayed in context. In order to differentiate between the own assigned tags and tags of other users, different colors are used.

For all three versions together, 97,772 tags for 11,306 artworks have been assigned by 182 users. For a closer discussion of the user vocabulary, we refer to [1]. Further analyses in this paper will focus on eight users that have assigned tags in all three versions. Together, these users have assigned 89,058 tags, that is 91,1% of all tags. As shown in Table 1, the eight users vary greatly in the amount of tags that they have provided in the different environments.

Table 1: users and the amount of annotated tags for the different tagging-environments.

Username	Total Tags	Tags V1	Tags V2	Tags V3
UserA	49,058	4,992	43,260	806
UserB	13,822	820	12,680	322
UserC	9,139	2,190	5,384	1.565
UserD	6,895	3,065	3,518	312
UserE	5,500	1,210	3,833	457
UserF	2,016	341	1.617	58
UserG	1,513	45	1,390	78
UserH	1,115	65	614	436
Total	89,058	12,728	72,296	4,034

5. RESULTS

5.1 Comparing User Motivation in Tagging Environments

We evaluate the three different tagging environments according to the heuristics of Körner et al. [8] [9], presented in Section 2. Figure 2 presents the four heuristics cumulated for V1, V2 and V3. Each score is normalized to a [0;1] interval.
The vocabulary size $vocab(u)$ is by far the greatest in V2. It actually improved for all eight users. Therefore it can be argued that the possibility to view already assigned tags encourages users to look at artworks more closely, and thereby annotate further details.
The average amount of tags per post $tpp(u)$ is lower for V2, where the users were able to view already assigned tags, than it is for V1 and V3. By viewing already assigned tags, the users were able to amend already existing tags, instead of assigning the complete set of tags all over again. Therefore the artworks are tagged more descriptive, as also stated by the tag-resource-ratio $trr(u)$, which increased for V2. The $trr(u)$ score clearly shows that artworks that were annotated in V2 have the highest amount of tags.
The orphan-ratio $orphan(u)$ is only slightly lower in V2, stating that users do not change the ratio of seldom assigned tags to the total amount of tags. A possible explanation might be that the users had already found most of their vo-

cabulary in V1, and therefore assign seldom used tags to only a few resources and frequently used tags to many resources. Since tags in V2 were added to already assigned tags, the orphan-ratio has is lowest score in V2. The scores of the heuristics for the explorARTorium (V3) are similar to the scores for V1.

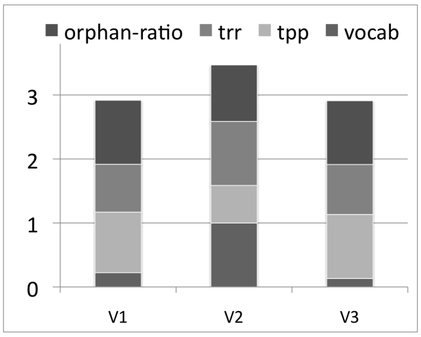

Figure 2: comparison of the different tagging environments according to its effect on the user's tagging motivation. All scores are normalized to an [0;1] interval.

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from descriptive to categorizing and vice-versa, we examined the tagging behaviour of all users that assigned tags in the three different tagging environments. Consequently, we applied the presented measures from Section 2 to the three tagging environments.
Figure 3 displays the scores of the vocabulary size $vocab(u)$, the tag/resource ratio $trr(u)$, the average tags per post $tpp(u)$ and the orphan-ratio $orphan(u)$ for V1, V2 and V3. In order to compare the relative distribution among the users, the scores were normalized to a [0;1] interval. The results indicate that there is no real change in user behaviour. Users who score high on the heuristics for V1, score even higher for V2 and V3. Users with a comparatively high tag/resource ratio, a high number of average tags per post and orphan-ratio, like UserA and UserE, are even encouraged to describe the artworks in more detail when already assigned tags are displayed. This indicates that the user motivation stays the same for all three environments. It can therefore be argued that users have different motivations for tagging, and that they did not change it due to the presentation of an artwork. The design of a tagging interface could only improve the total amount of assigned tags, but it was not possible to change the motivation of the users. This is in accordance with Strohmaier et al., [14] who compared different tagging platforms and argue that the tagging behaviour of users varies within the same platform (and therefore the same tagging environment).

As described in Section 3, the users have been able to rate artworks on a scale of zero to five, with zero being the lowest rating and five being the best rating. Previously introduced measures from Section 2 are applied on the average ratings of artworks. As shown in Figure 4 there is a correlation between the average rating of an artwork and the tagging behaviour of the users. Again, all scores have been normalized to an interval of [0;1]. The cumulated measurements score

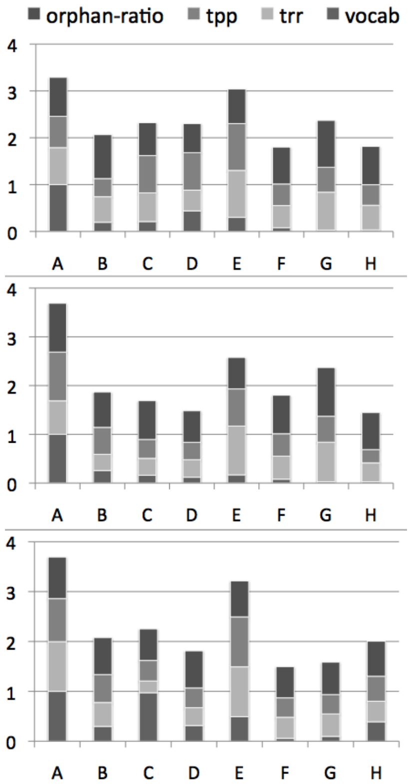

Figure 3: the user's tagging motivation in different the tagging environments. From top to bottom: V1, V2 and V3.

higher for artworks with higher ratings. The relatively high scores for artworks with an average rating of 0-0.5 might be due to the users dislike of them. Artworks that are disliked evoke feelings and further interest, and therefore might be tagged it in great detail. The vocabulary size reaches its peak at artworks with an average rating of 3-3.5. An explanation for this behaviour might be, that this group contains most artworks. Nevertheless, the vocabulary size is greater for artworks with a high rating compared to artworks with a low rating. The tag-resource-ratio $trr(u)$ again shows that the scores rise from an average rating of 3.5 to both sides, peaking at the highest rating. The $tpp(u)$-score rises from an rating of 1, again peaking for artworks with an average rating of 5. The orphan-ratio is very similar for all ratings. It can therefore be concluded, that the personal perception has a strong influence on the amount of tags users are willing to annotate, and might therefore influence the motivation of the users to rather describe or categorize artworks.

5.2 Identifying Categorizing / Descriptive Tags

In order to identify categorizing and descriptive tags in the folksonomy, we evaluated the introduced $atpp(t)$ heuristic from Section 3. Since the folksonomy was collected with three different tagging environments, it is difficult to compare them adequately, since the user's motivation might vary. Therefore only tags from an homogeneous environment are compared. Due to the amount of collected tags, we chose to analyse the tags from V1 together with the uninfluenced tags from V2. Consequently, we only apply

Table 2: top ten categorizing and descriptive tags according to the lowest / highest *atpp*-score.

Categorizing Tags						Descriptive Tags					
tag	$atpp(t)$	min	max	$users(t)$	$resources(t)$	tag	$atpp(t)$	min	max	$users(t)$	$resources(t)$
sistinechapel	4.83	1	12	4	12	relief	18.65	1	44	3	15
martyrdom	4.84	1	14	3	15	flies	17.76	1	44	3	16
oldman	4.86	1	10	4	15	bushes	17.22	1	38	3	33
fresco	4.92	1	15	7	20	lances	17.03	1	45	5	33
people	4.92	1	22	6	15	podium	16.94	2	37	3	14
michelangelo	5.06	2	12	6	10	gate	16.84	3	40	3	18
venice	5.09	1	9	8	20	birdcage	16.35	5	30	3	19
saint	5.14	1	14	3	14	gras	16.35	1	40	5	18
fresko	5.32	1	27	8	21	picknick	16.23	1	40	5	11
church	5.35	2	11	8	18	steps	16.20	1	63	5	101
portrait	5.36	1	25	21	390	stones	16.04	1	63	3	76

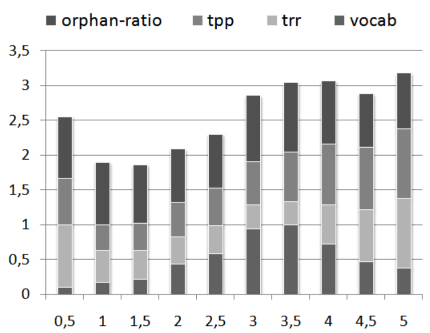

Figure 4: user's tagging motivation according to the average rating of an artwork.

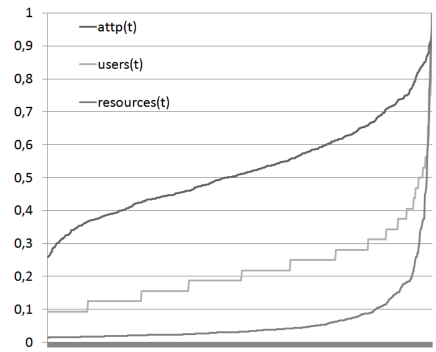

Figure 5: distribution of the $atpp(t)$ scores for all tags with $R_t > 10$ and $U_t \geq 3$.

the heuristic to tags that were collected while no additional tags where shown – either because the resource was not yet tagged or because the users did not actively demand to see them. To identify tags that can be objectively classified as categorizing, we only analyse tags that are annotated for more than ten resources ($R_t > 10$) by at least three different users ($U_t \geq 3$). Therefore 47,791 tags (637 distinct) for 5,891 paintings are considered in further analysis.
Figure 5 shows the distribution of the $atpp(t)$ heuristic, together with the distribution of the amount of users that annotated a tag $users(t)$, and the amount of resources that a tag was assigned for $resources(t)$. The x-axis represents a set of tags with $R_t > 10$ and $U_t \geq 3$ and the y-axis represents the scores, normalized to an interval of [0;1] in order to compare them adequately. For the $atpp(t)$ measure, low scores are considered to be categorizing and high numbers are considered to be descriptive. Due to the shape of the distribution, it can be assumed that the $atpp(t)$ heuristic distinguishes tags that are averagely annotated with only a few additional tags (categorizing) and a lot of additional tags (descriptive). The relative small amount of tags at the steep lower and upper end of the graph can be interpreted as categorizing, respectively descriptive tags. The other tags can be regarded as not sufficiently classified. The amount of users and resources follows shows a long-tail distribution.

The next step is to evaluate the tags themselves. We have therefore identified the tags with the ten highest and lowest $atpp(t)$ scores. Table 2 provides an overview of the ten tags with the lowest and highest $atpp(t)$ scores, together with the minimum and maximum amount of tags per post (min and max), the amount of users that annotated the tag $users(t)$, and the amount of resources that tags were assigned for $resources(t)$. According to Table 2, tags with the highest and lowest $atpp(t)$ score seem to be classified correctly. The categorizing tags represent locations, genres and the topic of artworks. Descriptive tags with a high $atpp(t)$ score annotate mostly details of artworks. The $atpp(t)$ heuristic seems to work rather correctly for the top/bottom fifteen percent of tags, an indication that is already given by the distribution shown in Figure 5.
By looking at the the maximum amount of tags per post (max), it can be observed that it is much lower for categorizing tags than it is for descriptive tags. It can therefore be argued whether the results are somewhat biased. Not all artworks allow the possibility to be annotated with the same amount of tags, since they just do not provide enough content or information. A fresco for example, which is represented as a whole picture, might make it difficult for a user to identify details. The same problem exists with portraits: after identifying the person together with some attributes, there might simply not be any more details left to tag. Whereas religious artworks or landscapes provide more information to tag than others. As shown in Figure 6, the average amount of tags per post varies greatly among the different themes. The x-axis represents the five most annotated themes of artworks for $R_t > 10$ and $U_t \geq 3$, while the y-axis represents the $atpp(t)$ scores for artworks of a specific theme. Religious paintings get $atpp(t)$ scores

between 4.5 and 19.5 (M=10.5; SD=2.9) and landscapes between 5.1 and 16.9 (M=9.7; SD=2.6). Portraits have the lowest $atpp(t)$ scores, with a minimum of 3.6 and a maximum of 12.4 (M=7.4; 1.9). Even though the results for each theme show that the $atpp(t)$ is able to distinguish tags among a single theme, it is difficult to compare tags of different themes. This becomes even more difficult when different epochs, schools or even artists are considered that created either abstract or realistic paintings.

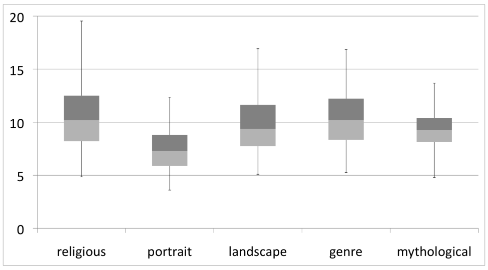

Figure 6: box plot of $atpp(t)$ scores for different themes with $R_t > 10$ and $U_t \geq 3$.

The same problem also exists for users. In order to compare the tagging behaviour of different users, we separately calculated the $atpp(t)$ scores for the eight users. Again, only tags with $R_t > 10$ and $U_t \geq 3$ are considered. According to the results shown in Figure 7, the $atpp(t)$ scores vary greatly among the users. UsersA [2.5;29] and UserE [1;27] get much higher $atpp(t)$ scores than other users, as UserA and UserE are probably describers as indicated in Figure 3. Therefore some kind of bias remains for the $atpp(t)$ heuristic, even though we somewhat reduce it by demanding that a tag should be assigned by at least three different users. Since users with a low range of $atpp(t)$ scores are probably motivated by categorization, the $atpp(t)$ heuristic accordingly distinguishes between categorizing and descriptive tags.

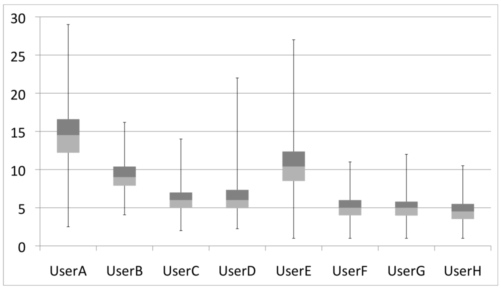

Figure 7: box plot of $atpp(t)$ scores in different personomies for tags with $R_t > 10$ and $U_t \geq 3$.

Finally, we analyse how fast users adopt the tags of other users and reuse them. Therefore it is measured how often a user had seen a specific tag until she herself assigned it for an artwork. We define tags with the lowest ten percent of $atpp(t)$ scores as categorizing tags and the highest ten percent as descriptive tags while only considering tags with $R_t > 10$ and $U_t \geq 3$. As seen in Figure 8, users adopt descriptive tags a lot faster than categorizing tags. The x-axis represents the amount of times a tag was seen, and the y-axis represents the number of distinct tags that were adopted and reused. Tags that are not classified as either descriptive or categorizing are not shown. Tags that were adopted after seeing them only once or twice are far more often descriptive than categorizing. Tags that were seen three of four times are slightly more often categorizing tags. The ratio is almost equal for tags that were adopted after seeing them at least five times. A possible explanation for this behaviour might be that descriptive tags are easier to identify, since they relate to the content of an artwork, whereas categorizing tags represent a more subjective interpretation of the artwork. This is in accordance with Körner et al. [9] who define categorizing tags as being more subjective than descriptive tags.

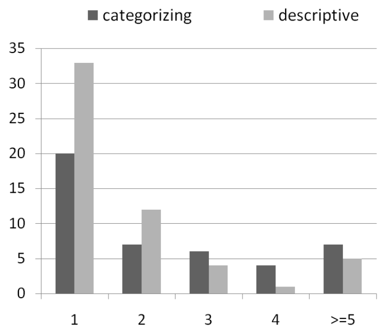

Figure 8: the amount of times a tag was seen until it was first assigned for a resource.

6. CONCLUSION

The works of Körner et al. [8] [9] and Stromaier et al. [14] distinguish two different kinds of users according to their motivation to assign tags: categorizers who categorize resources and describers who describe them. In this paper we use existing heuristics that distinguish the user's motivation, to evaluate our folksonomy. In order to differentiate tags in categorizing and descriptive, we introduce the $atpp(t)$ heuristic which calculates the average amount of tags the users assigned to a resource that is tagged with a specific tag t. This heuristic is based on the definition of Körner et al. [9], that describers can be identified by the amount of tags they assign to a resource. Describers tag a resource in great detail and therefore assign more tags to resources than categorizers, who annotate a resource very efficiently.

After outlining the ideas behind the art Web platform explorARTorium, which was used to collect about 97,772 tags for 11,306 artworks, we evaluated the folksonomy to identify how our tagging environment could be changed to collect more categorizing and descriptive tags. The results give insights that the design of tagging-interfaces for artworks did not necessarily change the motivation of the users to describe or categorize the artworks. The tags of a describer become even more descriptive when already assigned tags are displayed, but the motivation of users does not change just because the interface has changed. Describers stay describers and categorizers stay categorizers. This is in accordance with Strohmaier et al. [14] who argue that the motivation of the users varies among the same tagging-platform. We show that users adopt and reuse descriptive tags faster than they

do to categorizing tags – an indication that descriptive tags are more objective than categorizing tags, as also stated by Körner et al. [9]. Finally, we evaluated how well the $atpp(t)$ measure distinguishes between categorizing and descriptive tags. The results show that the $atpp(t)$ measure is able to correctly identify categorizing and descriptive tags. Nevertheless, the content of the tagged resource influences the score of the $atpp(t)$ measure, since not all artworks provide the same amount of visual detail to be tagged. Therefore, not all categorizing and descriptive tags are identified, but the results provide enough grounding to show that the identified categorizing tags can be used to derive a classification from the folksonomy, and that the identified descriptive tags can be used for searching and browsing.
Future work will focus on a better comparison of categorizing and descriptive tags among the personomies. The similarity of categorizing tags among users could provide indications about the users knowledge of art history. It would therefore be interesting to analyse how the presented heuristic can be used to estimate the expertise of users about certain themes, artists and epochs.

Acknowledgement

This work was funded by the FWF (Fonds zur Förderung der wissenschaftlichen Forschung / Austrian Science Fund), Project No. L602, "The Virtual 3D Social Experience Museum".

7. REFERENCES

[1] M. Arends, J. Froschauer, D. Goldfarb, and D. Merkl. Analysing User Generated Content Related to Art History. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, ACM, Graz, Austria, 2011.

[2] M. Arends, J. Froschauer, D. Goldfarb, D. Merkl, and M. Weingartner. Vermittlung kunstgeschichtlicher Inhalte durch die Kontextualisierung von Kunstwerken im explorARTorium. In *Konferenzband EVA 2011 Berlin: Elektronische & Medien Kunst, Kultur, Historie, die 18. Berliner Veranstaltung der Intaernationalen EVA-Serie Electronic Imaging & the Visual Arts*, Berlin, Germany, 2011.

[3] M. Arends, M. Weingartner, J. Froschauer, D. Goldfarb, and D. Merkl. Learning about Art History by Exploratory Search, Contextual View and Social Tags. In *Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies, IEEE*, Rome, Italy, 2012.

[4] L. Commare. Social Tagging als Methode zur Optimierung Kunsthistorischer Bilddatenbanken? Eine empirische Analyse des Artigo-Projekts. *Kunstgeschichte. Open Peer Reviewed Journal*, 2011.

[5] S. Golder and B. A. Huberman. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2):198–208, 2006.

[6] T. Hammond, T. Hannay, B. Lund, and J. Scott. Social bookmarking tools (i). *D-Lib Magazine*, 11(4), 2005.

[7] M. Heckner, M. Heilemann, and C. Wolff. Personal Information Management vs. Resource Sharing: Towards a Model of Information Behaviour in Social Tagging Systems. In *International AAAI Conference on Weblogs and Social Media (ICWSM)*, San Jose, CA, USA, 2009.

[8] C. Körner, D. Benz, A. Hotho, M. Strohmaier, and G. Stumme. Stop Thinking, Start Tagging: Tag Semantics Emerge from Collaborative Verbosity. In *Proceedings of the 19th International Conference on World Wide Web, ACM*, WWW '10, pages 521–530, Raleigh, North Carolina, USA, 2010.

[9] C. Körner, R. Kern, H. P. Grahsl, and M. Strohmaier. Of Categorizers and Describers: An Evaluation of Quantitative Measures for Tagging Motivation. In *Proceedings of the 21st ACM SIGWEB Conference on Hypertext and Hypermedia (HT 2010), ACM*, Toronto, Canada, 2010.

[10] C. Marlow, M. Naaman, D. Boyd, and M. Davis. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, HYPERTEXT '06, ACM, pages 31–40, Odense, Denmark, 2006.

[11] J. Oomen and L. Aroyo. Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. In *Proceedings of the 5th International Conference on Communities and Technologies*, Brisbane, Australia, 2011.

[12] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. Tagging, Communities, Vocabulary, Evolution. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, CSCW '06, ACM, pages 181–190, Banff, Alberta, Canada, 2006.

[13] M. Springer, B. Dulabahn, P. Michel, B. Natanson, D. Reser, D. Woodward, and H. Zinkham. For the Common Good: The Library of Congress Flickr Pilot Project: Report Summary, Library of Congress. Technical report, 2008.

[14] M. Strohmaier, C. Körner, and R. Kern. Why Do Users Tag? Detecting Users Motivation for Tagging in Social Tagging Systems. In *Proceedings of the International AAAI Conference on Weblogs and Social Media ICWSM2010*, Washington, DC, USA, 2010.

[15] J. Trant. Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept. In *Art Museums: Proof of Concept. New Review of Hypermedia and Multimedia*, Taylor & Francis, London, UK, 2006.

[16] J. Trant. Tagging, Folksonomy and Art Museums: Results of steve.museum's Research. In *Proceedings of Museums and the Web Conference 2009*, Archives Museum Informatics, Indianapolis, USA, 2009.

[17] J. Trant and B. Wyman. Investigating social tagging and folksonomy in museums with steve.museum. In *Proceedings of the Collaborative Web Tagging Workshop at WWW2006*, Edinburgh, Scotland, 2006.

[18] T. V. Wal. Explaining and Showing Broad and Narrow Folksonomies. Blog Post. `http://vanderwal.net/random/entrysel.php?blog=1635`, 2005.

[19] P. Walsh. The Web and the Unassailable Voice. In *Proceedings of Museums and the Web Conference*, Los Angeles, CA, USA, 1997.