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

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

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 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 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, users assign tags for artworks while competing against other users [4]. We created our own art Web platform, the explorARTorium, 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
lations 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 like or myfavorites. The proportion of factual, subjective and informative tags varies greatly for different resources and persononies.

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 a 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 explorARTorium. 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 \((u, T_u, r)\) 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 persononies of different sizes.

\[ \text{vocab}(u) = |T_u| \]

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.

\[ \text{trr}(u) = \frac{|T_u|}{|R_u|} \]

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.

\[ \text{tpp}(u) = \frac{\sum_r |T_{ur}|}{|R_u|} \]

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 \( T_{u,0} \) 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 orphan 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.

\[ \text{orphan}(u) = \frac{|T_{u,0}|}{|T_u|}, T_{u,0} = \{t||R(t)| \leq n\}, n = \left\lfloor \frac{R(t_{max})}{100} \right\rfloor \]
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.

\[ atpp(t) = \frac{\sum_{r \in R_t} \sum_{u \in U_t} |T_{ru}|}{\sum_{u \in U_t} |R_u|} \]

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 as on the feelings they had by recognising 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.cc.tuwien.ac.at/taggingtool
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.

<table>
<thead>
<tr>
<th>Username</th>
<th>Total Tags</th>
<th>Tags V1</th>
<th>Tags V2</th>
<th>Tags V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>UserA</td>
<td>49,058</td>
<td>4,992</td>
<td>43,260</td>
<td>806</td>
</tr>
<tr>
<td>UserB</td>
<td>13,822</td>
<td>820</td>
<td>12,680</td>
<td>322</td>
</tr>
<tr>
<td>UserC</td>
<td>9,139</td>
<td>2,190</td>
<td>5,384</td>
<td>1,565</td>
</tr>
<tr>
<td>UserD</td>
<td>6,895</td>
<td>3,065</td>
<td>3,518</td>
<td>312</td>
</tr>
<tr>
<td>UserE</td>
<td>5,500</td>
<td>1,210</td>
<td>3,833</td>
<td>457</td>
</tr>
<tr>
<td>UserF</td>
<td>2,016</td>
<td>341</td>
<td>1,617</td>
<td>58</td>
</tr>
<tr>
<td>UserG</td>
<td>1,513</td>
<td>45</td>
<td>1,390</td>
<td>78</td>
</tr>
<tr>
<td>UserH</td>
<td>1,115</td>
<td>65</td>
<td>614</td>
<td>436</td>
</tr>
<tr>
<td>Total</td>
<td>89,058</td>
<td>12,728</td>
<td>72,296</td>
<td>4,034</td>
</tr>
</tbody>
</table>

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 re-

Figure 2: comparison of the different tagging envi-

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

dressive 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 or-

dressive 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 or-

dressive 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 or-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

tressive 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 or-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

In order to evaluate whether different ways to present the artworks change the tagging motivation of a user from de-

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 com-
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 ($U_t ≥ 3$). 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 the five most annotated themes of artworks for the different themes. The x-axis represents the five most annotated themes of artworks for the different themes. The x-axis represents the five most annotated themes of artworks for the different themes. The x-axis represents the five most annotated themes of artworks for the different themes.

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

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

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

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 $U_t ≥ 3$, while the $y$-axis represents the $atpp(t)$ scores for artworks of a specific theme. Religious paintings get $atpp(t)$ scores

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

![Image](80x243 to 266x349)

**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.

![Image](83x573 to 264x672)

**Figure 7: box plot of \( atpp(t) \) scores in different folksonomies 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.

![Image](366x485 to 507x608)

**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 Strohmaier 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 exploreARTorium, 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 heuristics 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