Evaluating Tag-Based Information Access in Image Collections

Christoph Trattner* ctrattner@iicm.edu
Yi-ling Lin† yil54@pitt.edu
Denis Parra† dap89@pitt.edu
Zhen Yue† zhy18@pitt.edu
William Real‡ william.real@gmail.com
Peter Brusilovsky† peterb@pitt.edu

*Knowledge Management Institute, Graz University of Technology, Austria
† School of Information Sciences, University of Pittsburgh, USA
‡ Carnegie Museum of Art, Pittsburgh, USA

ABSTRACT

The availability of social tags has greatly enhanced access to information. Tag clouds have emerged as a new "social" way to find and visualize information, providing both one-click access to information and a snapshot of the "aboutness" of a tagged collection. A range of research projects explored and compared different tag artifacts for information access ranging from regular tag clouds to tag hierarchies. At the same time, there is a lack of user studies that compare the effectiveness of different types of tag-based browsing interfaces from the users point of view. This paper contributes to the research on tag-based information access by presenting a controlled user study that compared three types of tag-based interfaces on two recognized types of search tasks – lookup and exploratory search. Our results demonstrate that tag-based browsing interfaces significantly outperform traditional search interfaces in both performance and user satisfaction. At the same time, the differences between the two types of tag-based browsing interfaces explored in our study are not as clear.

Categories and Subject Descriptors

H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia - Navigation

General Terms

Human Factors

Keywords

Tag-Based Search Interfaces, Tag Navigation, Tagging Systems

1. INTRODUCTION

Social tags provide an easy and intuitive way to annotate, organize and retrieve resources from the Web. Promoted by several pioneering systems such as Delicious, Flickr, and CiteULike, social tagging has emerged as one of the most popular technologies of the modern Web. The value of tags was specifically advocated for image collections such as Flickr, where the presence of tags made images searchable and discoverable. While tags help to discover content even with a standard keyword-search, the most innovative feature of social tags was the ability to support browsing-based access to information through so-called "tag clouds". Effectively, tag clouds are a new "social" way to find and visualize information providing both: one-click access to information and a snapshot of the "aboutness" of a tagged collection. Not surprisingly, a large volume of research has been devoted to developing better approaches to construct and visualize tag clouds [5, 30, 18] as well as more advanced tag constructs such as clustered/classified tag clouds [23, 32, 2, 39, 16, 25] and tag hierarchies [10, 19, 34, 35].

The majority of research on tag clouds and hierarchies used an information- or network-theoretical approach to evaluate the quality of different tag constructs in terms of search and navigation while ignoring the user prospective. User studies comparing performance of users applying different tag-based browsing constructs in a set of realistic search tasks are rare. Moreover, there is a lack of user studies that compare the effectiveness of various tag constructs with simple search-based access to tagged collections. This paper attempts to bridge this gap by comparing several types of tag-based information access in a controlled user study. The study has been performed in the context of image search, where the presence of tags is known to be most valuable. To make the study more useful, we compared the performance of three types of tag-based information access interfaces in two commonly recognized types of search tasks – lookup search and exploratory search. The tag-based interfaces explored in the study include a search-based interface that plays the role of a baseline and two types of tag-based browsing interfaces: a regular browsing interface using traditional tag clouds and a faceted browsing interface using classified tag clouds. We selected the faceted tag cloud interface from among other advanced tag-based browsing approaches because our previous study [26] in the image search domain revealed that faceted search interfaces helped users to better explore large collections of images.

2. DATASET

As a dataset for our study, we utilized a collection of images from an archive belonging to the Carnegie Museum of Art in Pittsburgh, Pennsylvania. Overall, the collection contains more than

HT'12, June 25–28, 2012, Milwaukee, Wisconsin, USA.

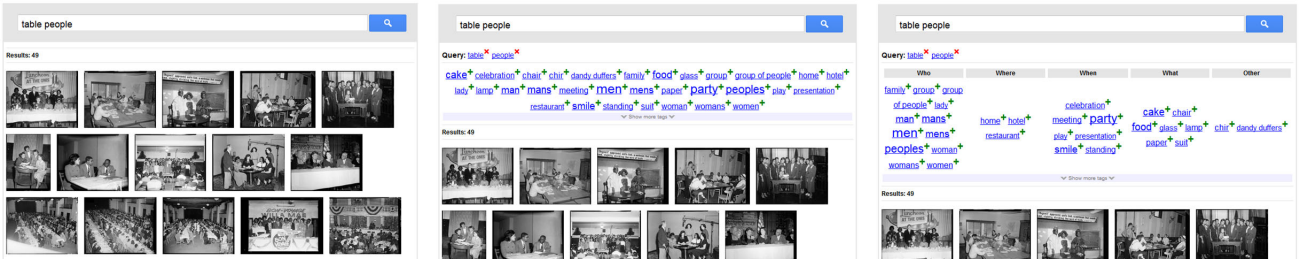

Figure 1: Screenshots of the three search interfaces - baseline (left), tag cloud (middle) and faceted tag cloud (right).

80,000 images taken by the famous local photographer, Charles Teenie Harris, who captured African-American life in Pittsburgh over a 40-year period. In our study, we used 1,986 of these images, of which 986 have been featured in a current exhibition at the Carnegie Museum of Art. The remaining 1000 images were included in this study as they provide a finer-grained overview of the entire collection. For the 1,986 images, we collected user tags using Amazon Mechanical Turk. Overall, the dataset provides 4,206 unique tags and 16,659 tag assignments applied by 97 users to the 1,986 images.

3. INTERFACES

For the purpose of our study, we implemented three tag-based interfaces to search the collection of Teenie Harris images – one standard "search box" interface and two interfaces that support both search and tag-based browsing. In the following section, we introduce these interfaces and their functionalities.

3.1 The Baseline (Search Only) Interface

As a baseline for our study (see Figure 1), we utilized a simple search box-based interface that offers the look and feel of well-known search engines. Similar to the Google, Yahoo! or Bing image search interfaces, we provide our users with a search box to submit a query, a thumbnail preview of the resulting images sorted by relevance and the functionality to click on the image in order to get a more detailed view of the image resource. The back-end of our search interface is built upon the OpenSource search engine Apache Lucene, which utilizes the tags of each image to create the search index.

3.2 The Tag Cloud Interface

The second interface explored in this paper is referred to as the tag cloud interface. As indicated by its name, this type of search interface extends the baseline search interface with the functionality of a traditional tag cloud. The alphabetically-ordered tag cloud provides the user with a topical overview of the search results and allows the user to search or browse images using the tags displayed in the cloud. This form of tag cloud is currently the most popular type of tag-based browsing in social tagging systems. To generate the tag cloud in this interface, we utilized a simple popularity-based tag cloud algorithm. For each query, we display the top N most frequent co-occurring tags to the user. This approach was shown to be one of the best choices to create a tag cloud from the perspective of tag-based search and browsing [37]. As the number of tags displayed in the tag cloud is an important factor which can negatively affect tag cloud-based search and navigation [33, 20], we also provide the functionality to increase or decrease the number of tags in the tag cloud to suit the user's needs. In Figure 1, a sample screenshot is presented to show how the tag cloud interface appears on the user's screen. As can be seen in the figure, the interface offers not only the functionality to click on a tag to issue a query, but also the possibility to expand the query by clicking the "+" sign in the tag cloud or shrink the query by utilizing the "x" sign in the query string beneath the search box. Currently, many popular tagging systems such as Delicious or BibSonomy offer similar approaches for query expansion or reduction to give the user a more flexible way to search and navigate in a tag-based information system.

3.3 The Faceted Tag Cloud Interface

The third interface developed for the study is referred to as a faceted tag cloud interface (see Figure 1). It can be considered as one of the most innovative tag-based search interfaces currently available. The interface was first introduced in 2009 by Yahoo! [32] in order to search for images in the social tagging system Flickr. Although there are very few implementations of this type of interface, there is a great deal of current research in this area [29, 38, 8, 7]. Similar to the tag cloud interface, this type of interface provides the user with the functionality to view the tags of the retrieved images in a visually appealing representation. However, contrary to the traditional tag cloud interface, where all tags appear in a tag cloud in an unstructured way, this type of interface classifies tags into several categories.

To decide which classification schema to utilize, we performed an extensive literature survey on currently available tag classification approaches [6, 29, 38, 8, 32, 11]. In the end, we selected a simplified form of the well-known "Editor's 5 Ws" approach that recognizes "Who" (people, groups or individuals), "Where" (location or places), "When" (time, activities or events), "What" (objects, food, animals or plants) and "Other" (unknown, not classified) classification schema. This schema was found to be effective in classifying tags in the image domain [32] as well as in our earlier user studies [26]. To classify our tags for this type of interface, we also used Amazon Mechanical Turk. The classification procedure itself was independent of image context as none of the currently available tag classification approaches take into account context information such as resource information, user information or other tags for the same or similar resources.

To ensure that the workers on Amazon Mechanical Turk (referred to as turkers) would classify our tags in a meaningful way, we provided them with detailed instructions of how to select those tags which fit into the one of the five given categories. The guidance included a sample screenshot of three different types of tags classified into one of the five categories and a detailed explanation of how to use these categories. Overall, three turkers were assigned to classify each particular tag. After the first classification round, we noted that 11% of tags were not classified as the turkers could not agree on which of the five given categories to use. Therefore, we decided to initiate a second classification round with an additional

Search Tasks	Search Task Descriptions
Lookup	Find the following picture!* 
Exploratory	1. Find at least 8 different types of stores/shops in Pittsburgh! Each type of store/shop should have at least two images from different locations, i.e. in total you will have to find at least 16 images.
	2. Pittsburgh is a city with many sport teams. Find at least 8 different sport activities! Each type of sport should be represented by at least two pictures. In total, you will have to get at a minimum of 16 pictures.
	3. Pittsburgh has a rich cultural heritage. There were many musicians who worked in Pittsburgh. Find at least 5 different types of music instruments which the musicians played in Pittsburgh. Each instrument needs 2 pictures and all pictures should be taken in different locations. In total, you will have to collect at least 10 pictures.

Table 1: Search tasks and descriptions (*= in the user study, only one image at a time was presented to the user).

six turkers (per tag) to increase the precision of our classification procedure. All in all, 22% of the tags were classified as "Who", 16% as "Where", 23% as "When", 34% as "What" and only 5% of the tags as "Other", which clearly out-performs current automatic tag classification approaches in terms of unclassifiable tags (represented as "Other" tags in our classification schema). We had 86 different turkers for the first classification round and 35 turkers for the second. The mean inter-rater agreement per tag over all turkers was substantial (75%).

In Figure 1, one can see a screenshot of how this type of interface appears on the user's screen. As with the tag cloud interface, users have the opportunity to issue a query by clicking on a tag, to expand a query by clicking on the "+" sign or shrink the query by utilizing the "x" sign in the query string beneath the search box. In addition, the faceted tag cloud can be expanded or collapsed as same as in the tag cloud interface.

4. USER STUDY DESIGN

To compare the three tag-based information access interfaces, we designed a within-subject study. In this design, each of our subjects evaluated the three different search interfaces during one study session. To determine when tag-based support is most effective; each interface was examined in the context of two kinds of search tasks, which are discussed in the following section.

4.1 Search Tasks

It has been shown that search task attributes affect the information seeking behavior of users [13, 36, 9]. The complexity, familiarity, clarity and difficulty of a search task influences how a person searches, browses and uses information systems [13, 17]. To account for the impact of these factors, our study separately evaluated the effectiveness of the three tag-based information access interfaces in the two primary types of search tasks known as lookup search and exploratory search.

As indicated by its name, lookup search is typically performed to find a specific information item in a document collection [27]. Lookup search tasks are considered to be relatively simple and most frequently involve using a traditional search interface (cf. [13, 36, 9]). More complicated search tasks – "beyond lookup"– are typically called exploratory search tasks [27, 9]. Exploratory search assumes that the user has some broader information need that cannot be simply met by a "relevant" information item (as in simple lookup search), but requires multiple searches interwoven with browsing and analysis of the retrieved information [26].

To study lookup search behavior, we created nine different lookup search tasks. All of these tasks were of a similar nature: the subject was given and the user was expected to find relevant images in the collection within a certain time limit. To account for the differences in difficulty [13, 36, 9], a variety of pictures were selected ranging from "easy" to "hard" to find. To classify images by difficulty, we calculated the mean search time for each image in the image collection based on lookup searches performed with Amazon Mechanical Turk. Then, we selected nine images ranging from "easy" to "hard" to find in the Teenie Harries image collection. In Table 1, the nine different images chosen for the user study are presented.

To study exploratory search behavior, we designed three exploratory search tasks as shown in Table 1. To ensure the balance between each type of user interface and also to capture the attribute of difficulty, we designed the exploratory search tasks carefully with a variety of additional search criteria and attributes. For instance, to capture balance with the faceted search interface, we tried to tune our search tasks to utilize as many facets as possible. We did that by asking the subjects to search for several different topics such as music, sports or shops as well as various search criteria such as different locations. To capture the property of familiarity with the search tasks, we asked our subjects in the post-questionnaire to rate their expertise level on the given topic or search item.

To be sure that our search tasks were meaningful, we performed several trial searches with Amazon Mechanical Turk and we conducted a pilot study.

4.2 The Process

As discussed previously, our subjects had to undertake two different kinds of search tasks using three different types of search interfaces within one user study session. During the study, each subject was assigned to perform nine different lookup and three different exploratory search tasks which were the same for the duration of the whole experiment. To counter the impact of fatigue and learning, the order in which the search tasks and system interfaces were used were rotated using a Latin square design. In addition to this, the lookup and the exploratory search tasks were randomized among all three interfaces to make sure that each of them was evaluated under different search interface conditions. The process of the user study was as follows:

1. Each participant was informed of the objective of the study, and asked to complete a consent form and a short questionnaire eliciting background information.
2. For each interface and task, a demonstration was given and the participant was given enough time to familiarize themselves with the interfaces and tasks. 3. For each interface, the user was given three lookup tasks and one exploratory search task.
   (a) Lookup task: An image was presented to the participant and a limit of 3 minutes (+30secs. for task reading)

Task	Measure	Baseline		Tag Cloud		Facet	
		All cases	Successful	All cases	Successful	All cases	Successful
Lookup	Cases	72	59	72	57	72	59
	Total Actions	9.01±.89	6.46±.67	8.58±.94	5.37±.56	8.68±.86	6.12±.63
	Search Time	77.35±7.35	54.19±5.31	75.38±8.03	44.37±4.48	77.67±7.8	52.17±5.32
Exploratory	Cases	24	23	24	20	24	22
	Total Actions	43.67±4.36	**42.17±4.27**	41.04±4.52	**33.50±3.37****	42.58±4.26	40.73±4.44
	Search Time	421.58±38.03	**413.48±38.81**	363.96±35.05	**312.4±30.74*****	378.33±33.46	356.91±32.8

Table 2: Descriptives (*mean*±*SE*) of total actions and search time by search and interface. Each statistic is calculated considering all cases and considering only successful search tasks (=significant at p<0.01 ; ***=significant at p<0.001).**

was given to complete the task. Afterwards, a post-search questionnaire was given to the subject to elicit disposition toward the system interface.

(b) Exploratory task: A description of the task was given to the participant and they were allotted a limit of 10 minutes (+1min. for task reading) to complete the task. A post-search questionnaire was presented as well.

3. A final questionnaire was given to the subject to assess the differences among the three search interfaces.
4. A series of open-ended questions were asked according to the observations made during the study.

4.3 Participants

Our study involved 24 participants (8 females, 16 males), who were recruited via email and flyers distributed throughout the University of Pittsburgh campus. The participants were from a variety of disciplines ranging from law to computer science. Four of them had earned a bachelor's degree, 16 a master's degree and four a PhD degree. The average age of the participants was 30.6 years (min=22, max = 61, SD=7.59 years). Almost all (except 2 participants) reported using computers for more than 5 hours a day. All participants (except two) rated their search engine skills as high and indicated using Google, Yahoo! or Bing frequently. A significant number (19) reported that they were familiar with tagging or used search tagging systems such as BibSonomy, Delicious or Flickr regularly. Four participants reported that they were familiar with the history of Pittsburgh, the rest of our subjects stated that they were not. On average, each user study session lasted 90 minutes.

5. RESULTS

In this section, we present the results of our user study. We start by comparing user performance with different search interfaces and follow with an extensive log analysis that describes how the interfaces were used. After that, we report the findings from our post and final questionnaires and report the participants' subjective opinions about these interfaces.

5.1 Performance Analysis

The main goal of this study was to compare user search performance for two types of search tasks (lookup and exploratory search) and with three different interfaces (with and without tag-based browsing support). To assess user performance, we examined search time and total number of interface actions [24] which are traditionally used in the study of search and browsing interfaces. Shorter search time and fewer actions should indicate a more efficient interface for image search.

While these two performance measures are known to be reliable, they do not allow us to clearly distinguish between several search conditions in the presence of many failed search attempts (i.e., cases where the subjects were not able to complete the task and were interrupted). Due to the presence of this cap, the time and actions spent on failed attempts flattens the overall differences, making different conditions look more similar than they are in reality. To avoid this problem, we separately measured user performance only on successful tasks. Given comparable success rates (as we observed in the study), user performance on successful tasks enables us to more easily distinguish between several conditions.

Table 2 provides a summary of performance data for our three interfaces and two kinds of search tasks. The table separately reports performance data for all tasks (including failed tasks with capped time) and that for successfully completed search tasks. As the data shows, the main difference in user performance is observed between the task types: exploratory search, as expected, required much more time and actions than lookup tasks. To discover significant performance differences among interfaces, we applied 2 x 3 ANOVA (analysis of variance). The analysis was done separately for search time and for the total number of interface actions as functions of search task and interface. We also separately evaluated data for all cases and for successful cases only. The analysis of successful cases data revealed significant differences between tag cloud and baseline interfaces in terms of search time, $p < .001$, and total actions, $p < .001$, under exploratory search. Likewise, we found a significant difference in the total number of interface interactions between faceted tag cloud and baseline (search only), $p = .037$. No significant differences were discovered for "the data for all cases". We also have not discovered any significant differences between the two kinds of tag-based browsing interfaces under all conditions.

Effect of familiarity and difficulty on performance. Prior research on exploratory search interfaces indicated that the value of advanced information access interfaces might depend not only on the type of task (i.e., lookup vs. exploratory search) but also on task difficulty [13] and user familiarity with the search topic [17]. In the context of our study, we registered some reasonable differences in user familiarity on a Likert scale(1-5) with the topics of the three exploratory search tasks (*M=3.125*, *SE*=.15056, *SD*=1.27751). In other words, it was possible to divide users into two groups for each task - those familiar with the task topic and those not. Moreover, as the study indicated, the level of difficulty in the three exploratory search tasks was considerably different between the one relatively easy task and the two more complicated tasks. These variations allowed us to perform a separate analysis that explored the combined effect of the interface, task difficulty, and task familiarity in the context of exploratory search. We ran a 3 x 3 ANOVA as a function of task difficulty and interface, and also controlling for the two levels of familiarity previously mentioned. As shown in Table 3, the analysis revealed a significant difference between tag cloud and baseline interfaces in search time for those users not familiar with the topic and at a medium level of task difficulty when considering all cases, $p = .014$, and when only considering successful cases, $p = .009$. No other significant differences were found. These results

Difficulty	Measure	Baseline		Tag Cloud		Facet	
		All cases	Successful	All cases	Successful	All cases	Successful
Hard	cases	6	6	7	4	6	4
	Total Actions	67.33±5.94	67.33±5.94	64.43±8.48	51.5±10.9	55.5±6.59	51.75±9.71
	Search Time	603.5±23.05	603.5±23.05	557.43±40.42	507.5±61.4	562.67±38.47	537.0±55.14
Medium	cases	3	3	4	4	3	3
	Total Actions	38.33±5.24	38.33±5.24	35.25±3.09	35.25±3.09	57.33±6.89	57.33±6.89
	Search Time	**494.67±148.17***	**494.67±148.17****	**285.75±16.95**	**285.75±16.95**	382.00±22.11	382.00±22.11
Easy	cases	5	5	5	5	6	6
	Total Actions	25.0±4.24	25.0±4.24	23.6±2.5	23.6±2.5	19.0±1.53	19.0±1.53
	Search Time	308.8±49.31	308.8±49.31	227.8±23.77	227.8±23.77	212.23±25.45	212.33±25.45

Table 3: Descriptives (*mean*±*SE*) of total actions and search time, by interface at different difficulty levels, when people are not familiar with the topics and under exploratory search tasks (*=significant at p<0.05).

indicate that the tag cloud interface provides the most significant impact in cases where tasks are more complicated and users are less familiar with the topic of the task.

A similar analysis of the impact of difficulty and familiarity was performed for the lookup search context, but we did not find significant differences between interfaces. However, the impact of difficulty and familiarity might be determined by the relatively low level of user task familiarity in this context. Based on the average of the ratings in the lookup search task (*M*=1.3611, *SE*=.08463, *SD*=.71809), our subjects were not as familiar with the images as they were in the exploratory task of the user study. Only two of them reported that they were familiar with the images due to the fact they found an image during a prior search session.

5.2 Looking Deeper: Log Analysis

Although the previous analysis reveals performance differences between interfaces and tasks, it does not show how different usage profiles were created for each of the interfaces and tasks. To look for these differences we performed extensive user log analysis on users' answering specific questions.

The first question was : *How different were usage profiles for different interfaces and tasks?* To build the usage profile, we distinguished several different interface actions: (1) *Search* (inserting a query in the search box); (2) *Click Tag* (issuing a query by clicking on a tag); (3) *Add Tag* (expanding the query with a tag by clicking the "+" sign); (4) *Remove Term* (removing a term from the query by clicking the "x" sign); (5) *Show More Tags* (clicking the show more tags button to increase the number of tags in the tag cloud): (6) *Show Fewer Tags* (clicking the show fewer tags button to reduce the number of tags in the tag cloud); (7) *Show More Results* (clicking the show more results button to increase the number of images in the result list); (8) *Click Image* (clicking on an specific image) and (9) *Total Actions*.

Table 4 presents usage profiles for different interfaces and search tasks. The most visible (albeit trivial) result is that the action *Search* is used more frequently in the baseline interface, $p = .006$. While the *Search* action is also used more frequently in the tag cloud than in the faceted tag cloud interface, this difference is not significant. Another interesting discovery is that the use of *Show More Results* is significantly higher in the baseline interface than in the tag cloud, $p = .015$. The corresponding difference between the baseline and the faceted tag cloud is close to significant at the acceptable level of $p = .055$. Since the use of *Show More Results* is evidence that the top results returned by the last search or tag browsing action were not satisfactory, we can argue that tag browsing was more successful at providing relevant results. We can speculate that this result stems from the tag browsing interface's ability to provide a snapshot of the "aboutness" of the collection, guiding the user to a more successful choice of a search term or tag. In addition, we found an intriguing difference between the tag cloud and the faceted tag cloud interfaces: the action *Add Tag*, which was used to narrow the results by adding tags to the query, was used significantly more frequently in the faceted interface than in the tag cloud interface, $p = .006$. The difference among interfaces in terms of the usage frequency of other actions (*Click Tag*, *Remove Term*, *Show More Tags*, *Show Less Tags*) was not significant. Table 4 also reports differences in the usage profile between lookup and exploratory search tasks. As we can see, the usage profile was considerably different for the two types of tasks. This emphasizes that lookup and exploratory search tasks are radically different from the user perspective. However, as users had different amounts of time available to complete lookup and exploratory search tasks, we compared percentages instead of the mean number of actions. However, to test for significant differences between these percentages, we run one chi-squared test per each action. As shown in Table 4, we found significant differences for the *Search* action, $p < .001$, the *Add Tag* action $p < .001$, the *Remove Term* Action, $p < .001$ and the *Show More Results* action $p < .001$. These indicate that people rely more on the search box, the *Add Tag* and *Remove Term* functionality, and skimming through the paginated results list in lookup tasks than in exploratory search tasks. The significant difference for *Click Image* action, $p < .001$, shows that people rely more on clicking images in exploratory search than in lookup search.

The second question that we attempted to answer was : *Does tag grouping by semantic category affect the usage of these categories?* As outlined in Section 3.3, we classified tags in our tag corpus into the following five dimensions: Who, Where, When, What and Other. The users in the faceted interface case were able to see which category each tag belonged. However, the users of both the search and regular tag cloud interfaces used the same terms in search and browsing, although without knowing to which category the issued query term or the clicked tag belonged. One could hypothesize that the tag usage profile (i.e., frequencies of using tags in different categories) may be affected by making the categories visible. Table 5 shows the proportion of query terms in each classification category as used by the study participants; each row presents percentages for each type of interface. We analyzed the significant difference in these percentages by running two chi-square goodness of fit tests. Considering overall tag usage, (i.e., aggregating lookup and exploratory search tasks), as well as setting the expected percentages of the tag categories to match those in the faceted tag cloud interface, we found them significantly different than those in the baseline interface ($\chi^2(4,548) = 46.092$, $p < .001$), and the percentages in the tag cloud interface ($\chi^2(4,683) = 58.612$, $p < .001$). This data provides evidence that explicit tag categorization does impact user behavior.

	Interface			Task			
	Baseline	Tag Cloud	Facet	Lookup	%	Exploratory.	%
Search	**9.24±.96****	**5.61±.82**	**4.81±.63**	3.89±.28	**44.45%*****	14.54±1.36	**34.27%**
Click Tag	.00	2.88±.46	2.92±.46	.94±.13	10.67%	4.92±.74	11.58%
Add Tag	.00	**.61±.14**	**1.25±.2****	.72±.11	**8.19%*****	.33±.12	**0.78%**
Remove Term	.00	.95±.18	1.40±.25	.62±.1	**7.08%*****	1.26±.3	**2.97%**
Show More Tags	.00	.17±.07	.11±.05	.07±.02	0.74%	.18±.09	0.42%
Show Less Tags	.00	.02±.01	.00	.01±.0	0.05%	.01±.01	0.03%
Show More Results	**1.78±.3****	**.86±.18**	1.01±.19	1.31±.17	**14.90%*****	.96±.2	**2.25%**
Click Image	6.66±1.1	5.59±.86	5.66±.87	1.22±.07	**13.90%**	20.22±.99	**47.65%*****
Total Actions	17.68±1.99	16.70±1.95	17.16±1.94	8.76±.52	100%	42.73±2.5	100%

Table 4: Summary of the *mean*±*SE* of actions based on each task session in the baseline, tag cloud, faceted tag cloud interfaces and means/percentages of actions based on each task session and interface for lookup and exploratory search tasks (=significant at p<0.01, ***=significant at p<0.001).**

	Lookup Task			Exploratory Task		
Question	Baseline	Tag Cloud	Facet	Baseline	Tag Cloud	Facet
1. Did the interface provide enough support for that task?	**2.88±.24**	**3.92±.15***	**4.04±.15****	**2.88±.21**	**4.21±.13***	**4.21±.13***
2. Were some of the interface features unnecessary for that task?	**1.33±.12**	**1.83±.18***	**1.92±.2***	**1.33±.12**	**1.54±.13***	**2.17±.23***
3. Were you confident in the system's ability to find relevant information on this topic?	-	-	-	**3.25±.22**	**3.92±.2****	**3.92±.18****
4. Did you find the tag cloud/faceted tag cloud helpful in finding relevant information?	-	3.79±.2	3.96±.18	-	4.13±.22	3.83±.21
5. Was it helpful to display the tags in different font sizes?	-	3.5±.23	3.54±.23	-	3.17±.27	3.38±.24
6. Was the + useful to add terms to the query?	-	3.77±.25	3.82±.27	-	4.05±.2	3.73±.27
7. Was the x helpful to remove terms from the query?	-	**4.09±.23****	**3.65±.26**	-	4.04±.17	4.04±.18
8. Did you find it distracting that some terms in the faceted tag cloud were not classified correctly?	-	-	2.33±.26	-	-	2.43±.25

Table 6: Response (*mean*±*SE*) to post questionnaire items (*=significant at p<0.05 ; **=significant at p<0.01 , scale 1-5, higher values indicate more agreement with the statement).

	Who	Where	When	What	Other
Baseline	9.9%	29.6%	11.7%	42.7%	6.2%
Tag Cloud	13%	28.8%	9.4%	43.2%	5.6%
Facet	16.2%	24%	18.9%	34.6%	6.3%

Table 5: Percentage of search actions in each type of semantic category by search interface.

5.3 Post-Task Questionnaires: Participants' Perceptions of the Interfaces

To better understand the participants' perceptions of each interface, we focus on analyzing user feedback about the different interfaces and their features. In the user study, the participants were asked to compute a post-task questionnaire after each of their search tasks was finished. By analyzing this questionnaire, we could assess the usefulness of each interface and see whether any significant differences could be found among the three interfaces and also between two search tasks (lookup vs. exploratory). Table 6 shows the average user rating for each question in the survey.

In Question 1 and 2, a 2 x 3 ANOVA was conducted on users' ratings in order to examine the effect of interface and search task. There is no significant interaction between interface and search task. For Question 1, a simple main effect analysis showed that there is a significant difference between the interfaces $F(2,46) = 30.113$, $p < .001$. Participants judged the support provided by the tag cloud interface significantly higher than that provided by the baseline, $p < .001$. They also rated the interface support of the faceted tag cloud interface significantly higher than that of the baseline, $p < .001$.

For Question 2, we also found a significant difference between the interfaces $F(1.406,32.332) = 11.097$, $p = .001$. Participants felt that the baseline interface had fewer "unnecessary features" than tag cloud, $p < .001$, and the faceted tag cloud, $p < .001$. However, the unnecessary features were a relatively trivial concern to the users of all three interfaces.

Question 3 specially asked about the exploratory search task : "How confident were the participants in the systems' ability to find relevant information". A 1-way ANOVA was used to test for performance differences among the three interfaces. We found a significant difference among the interfaces $F(2,46) = 5.412$, $p = .008$. The participants were significantly more confident in their ability to find relevant information with the tag cloud interface, $p = .015$, and the faceted tag cloud interface, $p = .037$, as compared to the baseline interface.

In Questions 4–7, we investigated the usefulness of various tag-related features. The 2 x 2 ANOVA as a function of interface (tag cloud and faceted tag cloud interfaces) and search task showed that the only significant difference within this group of questions "Was the x helpful to remove terms from the query", $F(1,20) = 6.450$, $p = .02$. The result indicated that users found this interface feature significantly more useful in the tag cloud than in the faceted tag cloud interface. No significant difference was found between the lookup and the exploratory search tasks in respect to Question 8.

5.4 Post Questionnaires: Participants' Interface Preferences and Comments

Another useful source of user feedback was a post questionnaire that was administered after each participant completed the entire study. This questionnaire offered us an opportunity to ask users for their opinions about three different interfaces. By this point in time, users had gained practical experience with both types of tasks and all three types of interfaces. As shown in Table 7, when asked a retrospective question "*Which one of the interfaces did you*

Question	Interface					
	Baseline	(freq.)	Tag Cloud	(freq.)	Facet	(freq.)
1. Which one of the interfaces did you like/prefer most?	4.2%	(1)	**54.2%**	**(13)**	41.7%	(10)
2. Which one of the interfaces would you prefer for lookup search?	4.2%	(1)	41.7%	(10)	**54.2%**	**(13)**
3. Which of the interfaces would you prefer for exploratory search?	-	(-)	41.%	(10)	**58.3%**	**(14)**
4. Which of the interfaces would you suggest the Carnegie Museum of Art?	-	(-)	41.%	(10)	**58.3%**	**(14)**

Table 7: Percentages and frequencies (=freq.) about final questionnaire items.

like/prefer most?", 54.2% (13) of subjects preferred the tag cloud interface, 41.7% (10) the faceted tag cloud interface, and only 4.2% (1) preferred the baseline search interface. This data correlates well with the users' actual performance on tasks. At the same time, user feedback differed on "forward looking" questions designed to assess user preferences in future situations (such as "*Which one of the interfaces would you prefer for lookup search?*"). For both tasks, the faceted tag cloud interface emerged as most preferred for future use. In addition, none of the users clearly preferred the baseline interface for exploratory search tasks. It is interesting that our subjects reported divergent results when they were asked about preferences in general and for each specific task.

Further, we found that 58.3% (14) of subjects favored the same interface for both past and future use while the other 41.7% (10) of subjects indicated a preference for a different interface when working on at least one type of tasks in the future. In particular, among the 10 subjects who reported changing preferences, one subject who favored the baseline (search only) interface in the prior tasks switched to the tag cloud interface for exploratory search tasks.

We believe that the most likely explanation for the difference in interface preferences between past and future tasks is the interface complexity. While the baseline search interface is very familiar to our subjects, both the tag-based browsing interfaces were rather novel. Moreover, while the subjects might have had at least some experience with using the traditional tag cloud interface, the faceted tag cloud was new to all of them. It is reasonable that a user's opinion of a more complex interface might be less favorable during their first attempts in using it. At the same time, armed with some experience, the users expressed stronger preferences for the use of more complex and powerful interfaces in the future. This might explain the difference in users' answers to the question "*which of the interfaces they would recommend for Carnegie Museum of Art*" (i.e., to professionals working with images): 58.3% (14) of our subjects recommended the faceted tag cloud interface while only 41.7% (10) of subjects recommended the tag cloud interface; no one recommended the baseline interface. This indicates that tag-based browsing interfaces, particularly the faceted tag cloud interface, were evaluated to be more powerful and more preferred for experienced users.

The data also showed that the main difference in users' perceptions is between the baseline and the two tag-based browsing interfaces.Tag-based interface was preferred almost unanimously for both previous and future situations. At the same time, the difference between the two tag-based browsing interfaces is much less pronounced: the traditional tag cloud interface appeared to be a bit simpler and more preferred during previous tasks (which correlates well with the performance data), while the faceted tag cloud was perceived as a bit more powerful and preferred for future tasks.

Further support for this assessment of users' subjective preferences across the three interfaces is provided by analyzing their explicit rating for each interface (see Table 8). On a Likert scale(1-5), the average rating for the baseline (search only) interface was 2.75, 4.17 for the tag cloud interface and 4.04 for faceted tag cloud interface. From these statistics, we can see that the baseline interface was rated significantly lower than the tag cloud interface, $p = .002$, and the faceted tag cloud interface, $p < .001$. However, there is no significant difference between the tag cloud and the faceted tag cloud interfaces.

5.5 Looking Deeper: Comment Analysis

To explain differences in users' perceptions of the different interfaces and their features, we examined verbatim comments provided in the post questionnaire. Below, these comments are grouped by the type of the interface preferred by the user:

5.5.1 Preferred Baseline (Search Only) Interface

According to the 24 participants, only 1 subject preferred the baseline search interface. The reason why the user chose this type of interface favorite was the following:

> "I liked the search box most, because everything else distracted me. For me it is not necessary to have tags, because I have everything in my mind!" – P20

This subject identified that the simplest interface is the best as it did not distract by adding elements to the interface.

5.5.2 Preferred Tag Cloud Interface

Thirteen subjects preferred the tag cloud interface. Based on the feedback from the interview and open-ended question on why they preferred a particular interface, our subjects attributed their preference for the tag cloud interface to it being more effective than the baseline interface. They also felt that it was easier to use than the faceted tag cloud interface.

> "The tag cloud provided more information than search only and was less complex than the facet search interface" – P4

> "I think the tag cloud interface was very helpful for exploratory search tasks and the faceted tags are a bit harder because I have to figure out what facet to look at" – P3

> "I like tag cloud because it gives me new ideas and it is easier to use" – P21

Sometimes, the poor categorization of tags in the faceted tag cloud interface accounted for why our subjects preferred the non-faceted interface. They either thought the category of the facet was of low quality or irrelevant to the task.

> "The facet did not seem to identify tags well" – P1

> "I would recommend the faceted interface only if tags are rich enough and categorized correctly, otherwise tag cloud is better" – P8

> "I think the categorization was not good, it was not relevant to the task" – P19

Some of the subjects preferred the tag cloud interface because they thought that the different font sizes in the tag clouds made more sense than the categorizing tags. Furthermore, some of them didn't pay attention to the category at all.

Question	Rating
1. Overall how would you rate the Search interface?	**2.75±.22**
2. Overall how would you rate the Tag Cloud interface?	**4.17±.13***
3. Overall how would you rate the Faceted Tag Cloud interface?	**4.04±.15***

Table 8: Response (*mean*±*SE*) to final questionnaire items (*=significant at p<0.05; higher values indicate more agreement with the statement).

"I did not look at the facets at all as I just looked at the terms" – P12

"Font size attracted my attention more than the facets" – P18

"The font size helped me to get most relevant information quickly" – P24

5.5.3 Preferred Faceted Tag Cloud Interface

Overall, we had 10 subjects who preferred the faceted tag cloud interface. The reason for this preference can be categorized into three aspects. First, they thought that the faceted tag cloud interface provided them with more functionality.

"I like faceted tag cloud because the interface provided me with the most functionality" – P6

"For difficult search task the facet is useful and for easy tasks you can just ignore the facet feature" – P7

"The Faceted tag cloud interface seems to be a smarter interface" – P13

Second, our subjects opined that the faceted tag cloud interface organized tags in more meaningful ways than the tag cloud interface.

"I prefer faceted tag cloud interface because it shows more tags in an organized way, so I could find more information faster" – P2

"It is easy to find the tags that I needed in faceted tag cloud" – P11

"I like faceted tag cloud interface, because the interface is clearer and I always know where to find the tag" – P15

The third aspect is that some of our subjects thought that the faceted tag cloud suggested more keywords to them. The interface also inspired them to think of additional relevant key terms.

"I like the faceted tag cloud because it suggest more query options than the tag cloud" – P5

"The faceted tag cloud made me think of more useful keywords than the tag cloud" – P21

6. RELATED WORK

Tagging systems such as Delicious, Flickr, and CiteULike have emerged as one of the most popular technologies of the modern Web era. Tagging behavior has been widely studied with regards to either the structure of tagging systems [15, 31], or qualitative insights about tagging behaviors across small collections [3, 12, 28]. The collective tagging behavior of users seems to offer a strong platform for summarizing and indicating content popularity to improve Web search [1].

In the computer-supported cooperative work (CSCW) domain, researchers have noted that tags could be utilized to offer search signals to others in the community. Several ranking algorithms have been investigated to improve search performance within the tagging space, such as SocialSimilarityRank [4], and FolkRank [21]. In the HCI community, Furnas et al. discovered the similarities in the cognitive processes between generation of search keywords and tags [14]. Kammerer et al. investigated how to apply relevance feedback about tags to indicate users' interests in various topics as well as to enable rapid exploration of the topic space [22]. Although CSCW and HCI both have provided different approaches to improve Web search, the focus of those studies was only on optimizing search ranking algorithms.

To understand how people use tags in reality and to what extent tag-based browsing constructs support users during their information seeking processes, we are interested in exploring the usage and efficiency of tag-based search interfaces. From an interface point of view, several interfaces have been explored. While tags are used to discover content in a traditional keyword-based search context, the innovative usage of social tags also supports browsing-based access to information. For instance, in [30], the authors investigated a visualization technique, a tag cloud, to display tags to support search performance. They applied various dimensions to construct tag clouds for use in information retrieval usage. They explored parameters of constructing tag cloud layouts including font size, quadrant and proximity-to-largest-word during a presentation period or an interpretative period. The study showed that the list ordered by frequency is better for categorizing.

Another tag-based browsing construct is clustered tag clouds [39], which utilizes SOMs for visualization. The proposed approach not only facilitates the discovery of relationships between tags and corresponding content, but also improves tag-based navigation by clustering relevant tags. A similar idea, classified tag clouds, studied by Yahoo! Labs [32] classified tags by utilizing facets such as Wordnet. Their approach enabled Flickr photo browsing through different facets. Their analysis showed that users could effectively deploy query recommendations to explore large sets of images annotated with tags. Other studies [19, 34] explored another advanced tag construct, tag hierarchy, for tag-based navigation. By utilizing a decentralized search framework [34], the authors found that there are significant differences among different approaches to tag hierarchy construction in terms of success rate and average path length.

Since our primary goal intent in this paper is to explore whether the tag-based browsing constructs could provide any additional value to tag-based search, we apply the most popular interface layout, a tag cloud, as our basic tag interface and compare it to a traditional search box interface. Furthermore, according to our previous study [26] on image search, where we discovered that facets help users in exploring a large collection of images, we also investigate a faceted tag cloud interface in this study [32].

A similar study conducted by Sinclair and Cardew-Hall investigated the usefulness of tag clouds in terms of information seeking by analyzing the usage of tag clouds in a traditional search interface [33]. They found that subjects prefer tag clouds when the search task is more general, but favor issuing search queries, when more specific information is needed. Contrary to their study, our work is based on the domain of images where typically no descriptive content (such as page-text or abstract information) is given. Furthermore, we study three separate tag-based interfaces to discover

the differences between a traditional search interface, a search interface enriched with tag clouds, and search interface extended with faceted tag clouds. In this setting, we can clearly identify *how people use* each interface and *how they perform*. To the best of our knowledge, this is the first work that compares multiple tag-based search interfaces.

7. DISCUSSION AND CONCLUSIONS

The main goal of the presented study was to perform a comparative user evaluation of tag-based browsing interfaces against simple search-based access to tagged collections. We compared user performance and feedback for three types of tag-based information access interfaces in the context of two recognized types of search tasks – lookup search and exploratory search. As expected, we obtained empirical evidence that the two tag-based browsing interfaces were superior to the baseline (search only) interface. At the same time, the analysis of objective data (performance and action profile) and of subjective data (questionnaires) produced slightly different results.

From the users' perspective, both tag-based browsing interfaces were perceived to be superior to the baseline. The users indicated that these interfaces provided significantly enhanced support for both types of user tasks and reported significantly higher levels of confidence that relevant information would be found. They also ranked both tag-based browsing interfaces significantly higher "overall" than the baseline interface.

From the performance and log analysis, significant differences were found for the traditional tag cloud interface when used in the exploratory search context. The tag cloud interface was found to be significantly more efficient in terms of both time and actions than the baseline interface. We also found that the tag cloud provided the most significant impact upon more difficult tasks and when the user was less familiar with the core topic of the task. A deeper analysis of user actions revealed another argument in favor of the tag cloud interface - with this interface, the "show more results" action was used significantly less often than in the baseline interface. This indicated that, with the tag cloud, the users were more likely to receive useful results at the top of the ranked list. None of these differences appeared to be significant for the faceted tag cloud; its objective performance was inferior to the performance of the traditional tag cloud. In addition, neither objective nor subjective data revealed any significant differences between the traditional tag cloud and the more advanced faceted tag cloud.

Why was the more advanced tag-based browsing interface less effective than the simpler tag-based browsing interface? Why was the faceted tag interface not a significant improvement over the baseline (search only) interface from a performance aspect? The post-session questionnaire provided some answers to these questions. This questionnaire asked users to select their "preferred" interface in light of two aspects : looking at performance in the past and looking forward to potential future uses of these interfaces. While the traditional tag cloud interface was preferred in previous tasks (which correlated with the objective performance data), the faceted tag cloud interface was the most popular for future use. It was also the top choice to be recommended to museum professionals. This was a strong indication that the faceted tag cloud interface was perceived as more powerful in the long run, but too difficult to use at first. This speculation is further confirmed by users' comments. In these comments, subjects stressed several aspects in which the faceted tag cloud interface was superior to the traditional cloud, yet indicated that it was harder to use at first. This data revealed that the faceted tag cloud interface should be assessed in a longer-term study, which would allow users to gain experience and become more proficient in operating with more sophisticated interfaces. We plan to explore this hypothesis in our future studies.

We also should acknowledge that the most noticeable differences observed in the study were not between the interfaces, but between the lookup and exploratory search tasks. Our data further confirmed that these two kinds of tasks are radically different. Exploratory search tasks are much harder; they consume more time and require more actions than lookup search tasks. Moreover, the very structure of user activities was very different between exploratory and lookup search. The occurrence of traditional search decreased considerably perhaps because it was much harder to find right keywords for the query. In contrast, almost 50% of user time in exploratory search context was spent on examining specific documents that were important to understand the domain and identify the most useful terms. These results correlate well with the previous research on exploratory search.

Finally, we should acknowledge a few limitations of our study. First, we focused on the query-to-image part of tag-based information access since it was the most different aspect among the three interfaces. The explicit presence of tags can also also enhance image-to-image navigation and further increase the value of tag-based browsing. Additional studies are required to determine the value of tags in this context. In addition, by the nature of our studies, we were unable to investigate one potential weakness of tag-based browsing in respect to classic search. All tag-based browsing interfaces require some considerable screen space for a tag cloud or other tag browsing artifact. This might reduce the space needed to show search results and decrease the effectiveness of tag-based browsing. In our studies, this effect was minimal: the study was performed on a regular desktop screen and search results were shown as thumbnails, which occupied relatively little space. We believe that, in this context, tag-based browsing interfaces were able to present a sufficient number of results despite the decreased presentation space. As the results shows, the Show More Results action was called upon significantly less frequently for the tag cloud. However, this might be of concern for those cases of mobile search with limited screen space as well as for different kinds of objects that require more space in the results presentation area. This is one of the reasons that we hesitate to generalize the observed results on tag-based information access to non-image resources. This is another aspect that requires additional investigation. We hope to explore some of these issues in our future work.

8. ACKNOWLEDGMENTS

This work is mainly supported by grants from the BMVIT – the Federal Ministry for Transport, Innovation and Technology (grant no. 829590) and the Marshalplan Funding Agency, to the first author. Part of this work is also supported by CONICYT – the National Commission of Scientific and Technological Research, government of the Republic of Chile, to the third author and NSF grant DUE-0840597 to the last author. The authors wish to thank the Carnegie Museum of Art for use of the Teenie Harris image collection.

9. REFERENCES

[1] P. , G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? *Proc. of the international conference on Web search and web data mining*, pages 195–206, 2008.

[2] A. Ammari and V. Zharkova. Combining tag cloud learning with svm classification to achieve intelligent search for relevant blog articles. In *In Proceedings of the 1st International Workshop on Mining Social Media*, 2009.

[3] L. B. Baltussen. *Barbarians versus gatekeepers ? Tagging as a way of defining the ememrgent living archive paradigm*. PhD thesis, 2010.

[4] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web - WWW '07*, page 501, New York, New York, USA, May 2007. ACM Press.

[5] S. Bateman, C. Gutwin, and M. Nacenta. Seeing things in the clouds: the effect of visual features on tag cloud selections. *Proc. of the nineteenth ACM conference on Hypertext and hypermedia*, pages 193–202, 2008.

[6] K. Bischoff, C. S. Firan, C. Kadar, W. Nejdl, and R. Paiu. Automatically identifying tag types. In *Proceedings of the 5th International Conference on Advanced Data Mining and Applications*, ADMA '09, pages 31–42, Berlin, Heidelberg, 2009. Springer-Verlag.

[7] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu. Can all tags be used for search? In *Proceedings of the 17th ACM conference on Information and knowledge management*, CIKM '08, pages 193–202, New York, NY, USA, 2008. ACM.

[8] D. Böhnstedt, L. Lehmann, C. Rensing, and R. Steinmetz. Automatic identification of tag types in a resource-based learning scenario. In *Proceedings of the 6th European conference on Technology enhanced learning: towards ubiquitous learning*, EC-TEL'11, pages 57–70, Berlin, Heidelberg, 2011. Springer-Verlag.

[9] K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Inf. Process. Manage.*, 31:191–213, March 1995.

[10] K. Candan, L. Di Caro, and M. Sapino. Creating tag hierarchies for effective navigation in social media. *SSM '08: Proc. of the 2008 ACM workshop on Search in social media*, pages 75–82, 2008.

[11] I. Cantador, I. Konstas, and J. M. Jose. Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1), 2011.

[12] E. H. Chi and T. Mytkowicz. Understanding the efficiency of social tagging systems using information theory. In *HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 81–88, New York, NY, USA, 2008. ACM.

[13] A. Diriye, A. Blandford, and A. Tombros. When is system support effective? In *Proceedings of the third symposium on Information interaction in context*, IIiX '10, pages 55–64, New York, NY, USA, 2010. ACM.

[14] G. W. Furnas, C. Fake, L. von Ahn, J. Schachter, S. Golder, K. Fox, M. Davis, C. Marlow, and M. Naaman. *Why do tagging systems work?* ACM Press, New York, New York, USA, Apr. 2006.

[15] S. A. Golder and B. A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.

[16] Y. Hassan-Montero and V. Herrero-Solana. Improving tag-clouds as visual information retrieval interfaces. In *InScit2006: International Conference on Multidisciplinary Information Sciences and Technologies*, 2006.

[17] M. A. Hearst. *Search User Interfaces*. Cambridge University Press, 1 edition, 2009.

[18] M. A. Hearst and D. Rosner. Tag clouds: Data analysis tool or social signaller? In *Proc. of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences, HICSS '08*, Washington, DC, USA, 2008. IEEE Computer Society.

[19] D. Helic and M. Strohmaier. Building directories for social tagging systems. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 525–534, New York, NY, USA, 2011. ACM.

[20] D. Helic, C. Trattner, M. Strohmaier, and K. Andrews. On the navigability of social tagging systems. In *Proc. of 2010 IEEE International Conference on Social Computing*, pages 161–168, Los Alamitos, CA, USA, 2010. IEEE Computer Society.

[21] A. Hotho, R. J, C. Schmitz, and G. Stumme. Information Retrieval in Folksonomies : Search and Ranking. *Data Engineering*, 4011:411–426, 2006.

[22] Y. Kammerer, R. Nairn, P. Pirolli, and E. H. Chi. Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 625–634, New York, NY, USA, 2009. ACM.

[23] O. Kaser and D. Lemire. Tag-Cloud Drawing: Algorithms for Cloud Visualization. *Proc. of Tagging and Metadata for Social Information Organization (WWW 2007)*, 2007.

[24] D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Found. Trends Inf. Retr.*, 3:1–224, January 2009.

[25] K. Knautz, S. Soubusta, and W. G. Stock. Tag clusters as information retrieval interfaces. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, HICSS '10, pages 1–10, Washington, DC, USA, 2010. IEEE Computer Society.

[26] Y. Lin, J.-W. Ahn, P. Brusilovsky, D. He, and W. Real. Imagesieve: Exploratory search of museum archives with named entity-based faceted browsing. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–10, Nov. 2010.

[27] G. Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49:41–46, April 2006.

[28] J. Oomen, L. B. Baltussen, Sander Limonard, M. Brinkerink, A. van Ees, L. Aroyo, J. Vervaart, K. Afsar, and Riste Gligoro. Emerging Practices in the Cultural Heritage Domain - Social Tagging of Audiovisual Heritage. In *Web Science10:Extending the Frontiers of Society On-Line*, Raleigh, NC, USA., 2010.

[29] S. Overell, B. Sigurbjörnsson, and R. van Zwol. Classifying tags using open content resources. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, pages 64–73, New York, NY, USA, 2009. ACM.

[30] A. W. Rivadeneira, D. M. Gruen, M. J. Muller, and D. R. Millen. Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 995–998, New York, NY, USA, 2007. ACM.

[31] 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*, volume 4 of *CSCW '06*, pages 181–190. ACM, ACM Press, 2006.

[32] B. Sigurbjörnsson and R. van Zwol. Tagexplorer: Faceted browsing of flickr photos. In *Technical Report, Yahoo! Research*, 2010.

[33] J. Sinclair and M. Cardew-Hall. The folksonomy tag cloud: when is it useful? *J. Inf. Sci.*, 34:15–29, February 2008.

[34] M. Strohmaier, D. Helic, D. Benz, C. Körner, and R. Kern. Evaluation of folksonomy induction algorithms. *ACM Trans. Intell. Syst. Technol.*, 2012.

[35] C. Trattner, C. Körner, and D. Helic. Enhancing the navigability of social tagging systems with tag taxonomies. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, i-KNOW '11, pages 18:1–18:8, New York, NY, USA, 2011. ACM.

[36] P. Vakkari, M. Pennanen, and S. Serola. Changes of search terms and tactics while writing a research proposal a longitudinal case study. *Inf. Process. Manage.*, 39:445–463, May 2003.

[37] P. Venetis, G. Koutrika, and H. Garcia-Molina. On the selection of tags for tag clouds. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 835–844, New York, NY, USA, 2011. ACM.

[38] C. Wartena. Automatic classification of social tags. In *Proceedings of the 14th European conference on Research and advanced technology for digital libraries*, ECDL'10, pages 176–183, Berlin, Heidelberg, 2010. Springer-Verlag.

[39] A. Zubiaga, A. P. García-Plaza, V. Fresno, and R. Martínez. Content-based clustering for tag cloud visualization. In *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, pages 316–319, Washington, DC, USA, 2009. IEEE Computer Society.