Tag-Based Information Access in Image Collections: Insights from Log and Eye-Gaze Analyses

Yi-Ling Lin[1] , Denis Parra[2], Christoph Trattner[3], and Peter Brusilovsky[4],

[1] Department of Management Information Systems, National Chengchi University, Taipei, Taiwan
yl_lin@nccu.edu.tw
[2] Computer Science Department, Pontificia Universidad Catolica de Chile, Santiago, Chile
dparra@ing.puc.cl
[3] Department of Information Science & Media Studies, University of Bergen, Norway
christoph.trattner@uib.no
[4] School of Information Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA
peterb@pitt.edu

Abstract. Tag clouds have been utilized as a "social" way to find and visualize information, providing both one-click access and a snapshot of the "aboutness" of a tagged collection. While many research projects have explored and compared various tag artifacts using information theory and simulations, fewer studies have been conducted to compare the effectiveness of different tag-based browsing interfaces from the user's point of view. This research aims to investigate how users utilize tags in image search context and to what extent different organizations of tag browsing interfaces are useful for image search. We conducted two experiments to explore user behavior and performance with three interfaces: two tag-enabled interfaces (the regular and faceted tag-clouds) and a baseline (search-only) interface. Our results demonstrate the value of tags in the image search context, the role of tags in the exploratory search, and the strengths of two kinds of tag organization explored in this paper.

Keywords: Eye tracking, tag navigation, tag search interfaces, tagging systems, user behavior, and user study.

1 Introduction

Social tags provide an easy and intuitive way to annotate, organize and retrieve information resources. Different from classic metadata, social tags are short free-form labels to describe different items, such as blogs (Technorati), images (Flickr and Google images), and research papers (CiteULike). Tags are powerful descriptors to navigate complex information spaces [1], [2]. Beyond the usefulness of social tags for discovering content with a standard keyword-search [3], their most salient feature is to support browsing-based information access through a visual artifact called "tag cloud" [4], [5]. Tag clouds allow users to find and visualize information offering both one-click access to information and a snapshot of the "aboutness" of a tagged collection. Not surprisingly, much research has been devoted to developing better approaches to construct and visualize tag clouds [6]–[8].

A great amount of the usability work on tags or tag clouds has applied an information- or network-theoretical approach to evaluate the quality of different constructs in terms of search and navigation. However, the user perspective was seldom investigated. Although some studies explored the *usefulness of tag clouds from the perspective of displaying formats* [9]–[14], some evaluated the usability in *browsing and search tasks* [4], [6], [15], and some focused on *tag enhancements for supporting realistic search tasks* [4], [8], [14], most of them investigated the usability issues by conducting user studies with a simple tag selection task or with theoretical measures without a real user study in a search environment. This paper attempts to bridge this gap by comparing three information access interfaces in the context of image search with two commonly recognized search tasks – lookup search and exploratory search.

To discover how users utilize tags and how useful are different organizations of tag browsing interfaces in image search, we conducted two experiments to compare user behavior and performance with three interfaces (e,g, search-only, regular tag cloud and faceted tag cloud interfaces) for image search and exploration. The first experiment focused on revealing performance and action-level differences between different interfaces and search tasks. The second experiment employed an eye-tracking technology to reveal the sources of these differences.

To drive our research, the following two research questions were posed:

- RQ1: Are there any observable performance differences between the three investigated interfaces in the context of the two different image search tasks?
- RQ2: And if yes, which specific patterns may lead to these differences?

2 Research Background

Social tags are collaborative practices performed by internet users to annotate and share sets of online digital resources [16], [17]. According to the social Web's general principle of sharing and participating, social tags quickly established themselves as one of the most popular Web 2.0 technologies for transforming the static Web into a participatory information space [2]. This technology has demonstrated its value for information access in many different domains. For further tag history, tagging process, and accessing information with tags, interested readers may refer to Gupta's overview article [18] and Bullock et al.'s chapter [19] in the newly published social information access book [20].

While scholars have devoted efforts on investigating the social tagging process [21] and tagging motivation [22], [23], using tags to organize content has also caught scholars' attention from various aspects, such as tag ranking and selection [5], [24]–[26], tag recommendation [27]–[29], tag cloud construction [26], [30]–[33], tag visualization [34], [35] and tag applications to improve search and navigation[5], [15], [36]–[38]. Tag clouds have become one of the most common approaches to describe content, organize given resources and further enhance resource search, navigation, and recommendations [5], [39]. To address on the research goal of this work, we hereby review related work of tag clouds on user experience.

Among studies regarding tag clouds, one research direction focuses on assessing tag cloud display formats from the information finding perspective. Rivadeneira et al. [10] investigated the recognition of single tags in four tag-cloud layouts including alphabetical, sequential-frequency, spatially packed, and list-frequency with 11 participants. Their result did not show a significant effect of layouts in recognition of tags but subjects performed significantly better in category recognition with a tag cloud using a simple vertical list layout. Halvey and Keane [9] discovered the effect of tag cloud and list arrangements with different property settings e.g. alphabetization and font size. While no differences were found between layouts, they confirmed that alphabetization and font size variation significantly affect finding tags. Another comparative study conducted by Lohmann et al. [12] examined sequential, clustered, circular tag cloud layouts and their result shows that the design of layouts affects the performance of different information seeking tasks. Bateman et al. [13] examined nine visual properties of tag clouds. They concluded that font size and font weight have more significant impacts on finding a tag than other features; but when too many visual properties are applied at once, there was no distinct feature standing out over the others. Schrammel et al. [40] explored the effect of semantic arrangement of tags in tag clouds. Although the semantical clustered tag cloud can improve search performance for a specific search task compared to random arrangements, the alphabetic layouts still outperformed the semantic one. Bar-Ilan et al. [14] compared the effectiveness of four different tag-enabled interfaces (i.e. a search box interface, a faceted search box interface, a tag cloud interface, and an ontology-based tag search interface) by conducting a user study. The paper reported that users with the ontology interface achieved the highest recall on average for seven out of the ten tasks; however, users were more satisfied with the search box interface than the cloud or the ontology interface.

Past studies have also proved the value of tag clouds, particularly on search and browsing tasks. Kuo et al. [15] analyzed the utility of tag clouds for the summarization of search results from queries over a biomedical literature database with 20 participants. A tag cloud is considered as a more enjoyable result presentation of descriptive information than a standard result list; however, tag clouds performed significantly worse when presenting relationships between biomedical concepts. Another study [4] investigated the usefulness of tag clouds for different information seeking tasks. They found that participants preferred tag clouds for browsing scenarios, especially when the search task is more general, but tended to issue search queries instead when more specific information is needed. Such result indicates that tag clouds support users to explore unknown or unfamiliar items. Koutrika et al. [8] introduced a framework that generates data clouds from search results through the process of an entity-based approach. They investigated several algorithms for the extraction of data-clouds to improve the search process and examined the effectiveness of their proposed approaches for search and browse interfaces. Microsoft research team introduced SparkClouds by integrating sparklines [41] with typical tag cloud features to convey trends between multiple tag clouds [6].

Another research direction focuses on tag enhancement for improving search support. Bischoff et al. [42] demonstrated that tag classification scheme is useful for improving search. Different types of tags can provide additional information valuable for different source research. Böhnstedt et al. [43] introduced the semantic tag types as an extension of the tagging systems without losing their simplicity and accessibility. While faceted browsing interfaces have been shown to be an attractive and powerful alternative to "text box" search in situations when item metadata are available, such as Flamenco [44] or Relation Browser [45], scholars of tag related work also attempted to discover faceted tag representations. Collins et al. [7] introduced a similar faceted tag clouds, called parallel tag clouds, a visualization way for different tag usages among facets. Yahoo! [46] also demonstrated multiple tag clouds by the faceted tag-cloud interface to search images in the social tagging system, Flickr. Trattner et al. [38] conducted a user study to investigate the differences between a traditional tag-cloud and an advanced faceted tag cloud interface. Lin et al. [47] proposed a dual-perspective navigation framework (DPNF), that utilizes the mechanisms of faceted browsing and tag-based navigation to offer a seamless interaction between subject headings and social tags of image collections. Sciascio et al. [48] introduced a social exploratory search in a user-controlled way that allows users to express information needs based on their preferences.

Among all kinds of tag applications, tags are especially important for image collections on account of their limited textual information. To assist image search, classical image metadata (e.g. subject headings in the museum context) requires significant manual generation effort and suffers from the classic indexer-user mismatch problem [47]. Our previous work [38] has proven that tags are useful with simple and faceted tag-based interfaces are useful for search support, amplifying users confidence of finding more relevant information. This study extends the previous research trend in three distinct ways. First, we developed three different realistic search interfaces that are comparable to each other. In contrast to previous work comparing either a search-box based interface with a tag cloud only or other tag-constructs with each other, we assessed the value of the most realistic design - a tag cloud as an additive to a search box. Second, our study explores the value of tags in both look-up and exploratory search tasks with different levels of difficulty. Most of the previous work applied tag-selection, simulations or simple exploratory search tasks to assess the value of tags for the purpose of search and navigation. Third, a detailed log-based action analysis and an eye tracking study were conducted to obtain a deeper understanding of the usefulness of tags in an image search context. Our work is distinct by the combination of its scale, the use of a broader set of tasks, and the application of eye-tracking to uncover the mechanism of tag use and the reasons for the observed performance differences. Altogether, we believe that our work represents the most complete effort to date to investigate the usefulness of tags and tag clouds in the image search context.

3. Methodology

In this section, we present the details of the methodology of our research including the image dataset, the investigated interfaces, the search tasks, the study participants, the user study procedure and the statistical analysis we have applied in this study.

3.1 Dataset

This study utilized an image collection from the Carnegie Museum of Art in Pittsburgh, Pennsylvania. The collection contains images taken by the local photographer, Charles "Teenie" Harris, who captured African-American life in Pittsburgh over a 40-year period. We used 1,986 of these images, of which 986 were selected by the curators for the 2011 exhibition at the Carnegie Museum of Art and the remaining 1000 images were sampled by us to provide a more balanced representation of the entire collection. Since the original collection doesn't have any tags, we collected and classified tags for the selected 1,986 images using Amazon Mechanical Turk. While collecting the tags, we developed two tagging interfaces with image descriptions and without image descriptions on Amazon Mechanical Turk. For the further detail of our tag-collection process, interested readers can refer the third section of [49]. We adopted the tags without description in this current study. The dataset comprises 4,206 unique tags and 16,659 tag assignments applied by 97 users to the 1,986 images. To classify our tags into facets, we also used Amazon Mechanical Turk. The detailed explanation of how to classify tags into facets using Amazon Mechanical Turk can be found in [38].

3.2 Interfaces

We implemented three interfaces – a baseline "search box" interface with no tags, a traditional and a faceted tag-cloud enabled interfaces. The two tag-cloud interfaces support both search and tag-based browsing. Each interface includes two main "pages": the *result* and *detail* pages (Figure 1). The result page is the place where users issue a query (by typing in a search box or clicking on a tag) and check the list of returned images. In the tag-cloud interface, the result screen shows a traditional tag cloud that organizes tags for the returned images, while a faceted tag-cloud interface arranges tags for the returned images into several separate facets. The detail page is displayed when users clicked on a specific image. It shows the enlarged image and, for tag-based interfaces, also displays tags associated with this image as a traditional cloud (Figure 1, row T) or tag-based cloud (Figure 1, row F) correspondingly.

3.3 Search Tasks

Search task attribute has been proven to be an important impact factor on users' information seeking behavior [48], [50]–[52]. Various attributes like complexity, familiarity, clarity and difficulty of a search task influence how a person searches, browses and interacts with information systems [51]–[53]. To account for the impact of search task attributes, we investigated the value of tag-based image access in the two most common search tasks: lookup search and exploratory search.

Lookup search is typically performed to find a specific item in a document collection [54] and frequently associated with the traditional search interface [50], [51]. According to Gegenfurtner, Lehtinen, and Säljö [55] who distinguished four levels of task complexity in eye-tracking research, lookup search tasks can be considered as the

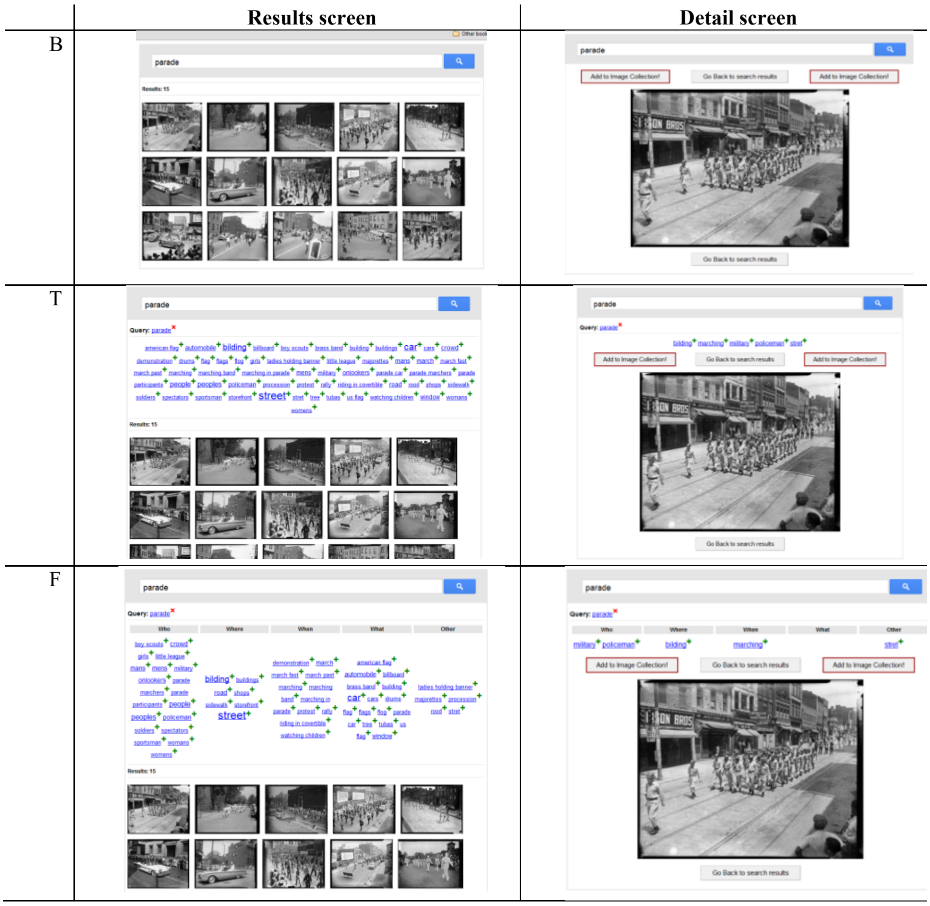

Fig. 1. The results and detail pages on three interfaces (baseline, tag-cloud, and faceted tag-cloud in rows). B denotes the baseline interface, T denotes the tag-cloud interface and F denotes the faceted tag-cloud interface.

combination of a viewing task and a detection task. We selected nine different images from the 1,986 image set as lookup tasks (Table 1). The subject was randomly assigned a target image and was asked to find it in the collection within a certain time limit. To account for the differences in difficulty [50], [51], we performed a lookup experiment on Amazon Mechanical Turk for each image in the image collection and calculated the mean search time for each image to define each image's difficulty. At the end, we selected nine images ranging from "easy" to "hard" to find in the image collection.

Table 1. Search tasks and descriptions.

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

*=only one image was presented to the user at one time

Exploratory search tasks require more complicated search behaviors [50], [54]. This kind of search assumes that users have some broader information need that cannot be simply met by one "relevant" item and requires multiple searches interwoven with browsing and analyses of the retrieved information[56]. With respect to the four levels of task complexity in eye-tracking research identified in [55], exploratory search tasks refer to the decision task and problem-solving task. We designed three exploratory search tasks described in the second row of Table 1. To ensure a balance between each type of user interface and also to control for difficulty, we designed the exploratory search tasks carefully with a variety of additional search criteria and attributes. We tried to tune our search tasks to utilize as many facets as possible. To capture the possible impact of user familiarity with the search tasks (i.e., Task 2 should be easier for a sports fan), we asked our subjects in the post-questionnaire to rate their expertise level on the given topic or search item. We performed several trial searches with Amazon Mechanical Turk and conducted a pilot study to ensure that our search tasks were meaningful.

3.4 Study Participants & Procedure

We conducted two user studies at the usability laboratory located at the University of Pittsburgh's School of Information Sciences. The first study included 24 student participants (8 females and 16 males) from a variety of disciplines raging from law to computer science. Their average age was 30.6 years old (min = 22, max = 61, SD = 7.59 years). Four participants reported that they were familiar with the history of Pittsburgh while the others stated that they were not. This study employed detailed logs recording as the principal data collection approach. Partial results of this study were reported in [38]. The second study included nine student participants (4 females and 5 males) from different disciplines with 20/20 vision. The age ranged from 25-35 (min = 20, max = 36, SD = 4.87 years). None of them was familiar with the history of Pittsburgh. We used an eye-tracking device for reveal aspects of user information processing that are not registered by the usual logs.

The 24 participants in the first study performed the search tasks on a regular desktop computer. The nine participants in the second study performed the tasks with a 17" LCD monitor integrating a Tobii 1750 eye tracker that provides a reasonable spatial resolution (spatial error is below 0.1 degrees of visual angle) and temporal resolution (20 ms sampling interval). A screen stimulus has a resolution of 1280 x 1024 pixels. Tobii 1750 software framework collected raw gaze data from dynamic stimuli with scene tool in ClearView software. ClearView maps a series of raw coordinates to a single fixation when the coordinates stay with a minimum threshold of 100 ms within a sphere of a fixation radius 30 pixels. We integrated eye-tracking data files obtained from ClearView for analysis of each task within each participant. The overall experiment was performed in a consistent way in the rounds with and without the eye tracker.

The procedure and the tasks conducted in both studies were identical. Each participant had to complete three lookup search tasks and one exploratory task using each of the three explored interfaces. To support users' work on search tasks, the interfaces shown in Figure 1 were augmented with a *working panel* including help area, task description, and a box for collecting search results. In addition, the interfaces in the eye-tracking study were slightly adjusted to avoid scrolling. The location of the task panel and the eye-tracking adjustment are explained in Figures 6 and 7.

In total, each subject was expected to complete the same nine lookup tasks and three exploratory tasks (Table 1). To account for the impact of fatigue and learning, the order in which the search tasks and system interfaces were used in the session were rotated using a Latin square design. In addition, the allocations of tasks to interfaces were randomized to ensure collection of sufficient data for each task-interface pair. The detailed process is listed as following:

(1) Each participant was informed of the objective of the study, and asked to complete a consent form and short questionnaire eliciting background information.

(2) A demonstration of the interfaces and tasks was given and the participant was given sufficient time (approximately 15 mins) to familiarize with the interfaces and tasks.

(3) For each interface, the user was given three lookup tasks and one exploratory search task. The description was displayed in the task area at the beginning of each task. To complete the task, the participant had to add the target images (one for a lookup task and a required number for an exploratory task) to the collection box. A button to add an image to the collection box was provided on each image detail page. A limit of 3 minutes (+30 sec. for reading) was given to complete a lookup task and of 10 minutes (+1 min. for reading) to complete an exploratory task. A post-search questionnaire was presented after the completion of each task type with each interface.

3.5 Statistical Analysis

The goal of our study was to investigate how well each of the three image search interfaces can support users in performing both lookup and exploratory search tasks. We assumed that a more efficient interface should allow users to complete the tasks faster and with fewer actions. According to the past research evidence [47], [56], we hypothesized that both tag-enabled interfaces will be more efficient than the baseline search interface in terms of both time and actions and for both kinds of tasks. We also hypothesized that the faceted tag cloud will be more efficient than the regular tag cloud for both kinds of tasks.

To assess these hypotheses, we performed click and time-based performance analysis with Multivariate Analysis of Variance (MANOVA) since we have multiple, dependent variables. The results of this analysis are presented in the next section. While the analysis confirmed part of the original hypotheses, it also revealed differences that were not matching our expectations. To examine the roots of the observed performance differences, we performed two additional rounds of analysis – a fine-grain log-based pattern analysis, and an eye-tracking pattern analysis. Among the three rounds of analysis presented below, the first two rounds are based on the log data collected from the first study, while the third one used eye-tracking data from our second study.

4. Results

This section presents the results of our two studies in the context of our research questions. The section 4.1 attempts to answer RQ1 from the performance perspective and the sections 4.2-4.3 starting specific patterns analyses to answer RQ2.

4.1. Log-based Performance Analysis

To assess the performance differences between the interfaces, we examined and compared search time and the total number of interface-level click actions for each condition. These actions include: clicking on the search button after typing a query, clicking on images to see their details, clicking on show more results, and clicking on tags (for faceted and tag-cloud interfaces). During the study, we had the presence of failed user sessions, i.e., cases where the subjects were not able to complete the task within required time limits. To simplify the analysis and discussion, we removed failed user sessions.

Table 2 shows the very distinct range of mean time and total actions on both tasks. While the average number of actions needed to complete each lookup task is between 5-6 actions per user and the time needed is from 44 to 54 seconds, for exploratory tasks the users require from 312 to 413 seconds and between 33 and 42 actions in average to complete each task. Given these results, the different nature and research design of our lookup and exploratory search tasks, hereinafter we analyzed them separately.

Table 2. Descriptive statistics (mean, S.E.) of total actions and search time by task and interface.

	Lookup			Exploratory		
	#	Average number of actions	Total time	#	Average number of actions	Total time
B	59	6.46±.67	54.19±5.31	23	42.17±4.27	413.48±38.8
T	57	5.37±.56	44.37±4.48	20	33.50±3.37**	312.4±30.74***
F	59	6.12±.63	52.17±5.32	22	40.73±4.44*	356.91±32.8

Two kinds of average data are calculated and checked for significance: considering lookup (left) and exploratory (right) task (*=significant at $p < .05$, **=significant at $p < .01$; ***=significant at $p < .001$). B denotes the baseline interface, T denotes the tag-cloud interface, F denotes as the faceted tag-cloud interface and # denotes the number of subjects

Table 3. Descriptive statistics of exploratory subtasks (Mean±SE).

Variable\ Task	Music	Sports	Stores
Search time	232.875±20.82*^	352.360±20.93*§	550.905±25.05^§
Total actions	21.625±2.29*^	41.874±2.3*§	57.314±2.75^§

All the pairwise comparisons are significant at the alpha < .001 level after Bonferroni correction. Symbols show the pairwise significance tests conducted between tasks: * music vs. sports, ^ music vs. stores, § sports vs. stores.

We first checked MANOVA assumptions, which are similar to those of univariate ANOVA. We tested the assumption of normality with Shapiro-Wilk tests, and in the case of exploratory tasks we found no deviations from normality (i.e., $p < .05$) for total actions, and for total time only in two out of the nine cells in the baseline interface with the medium difficulty ($p = .001$) and the faceted interface with the high difficulty ($p = .002$). However, in the case of the lookup search we found large deviations from normality. Eight out of nine cells significantly departed from normality in the case of total actions, and the same result was observed for total time, so we only proceeded with the MANOVA on the exploratory search data. With the respect to the assumption of homogeneity of variance, a Brown-Forsythe test was neither significant for total time $F(8, 56) = 1.088$, $p = .385$, nor for total actions $F(8, 56) = 1.728$, $p = .112$.

Another important factor to consider is the difficulty of the task. Past user studies of exploratory search interfaces revealed that while looking comparable at the design stage, tasks assigned to users might considerably differ by their effective difficulty [57], [58]. Similar results have been reported in eye-tracking studies. For example, Steichen and colleagues [59] investigated how to use gaze patterns to predict user abilities and task characteristics. To explore the potential impact of task difficulty on the user's work with different information access interfaces, we examined the total user performance (actions and time) between the three exploratory search tasks. Table 3 demonstrates that despite our attempts to make the tasks comparable, their effective difficulty was significantly different in time and actions. In this context, we defined the subtask on finding images of music instruments as of low difficulty, the subtask of finding sports' images as of medium difficulty, and the subtask of finding stores' images as of high difficulty.

Finally, using the data of the exploratory search task we conducted a 3x3 between-subjects MANOVA on total actions and total time as a function of interface and difficulty. Using Pillai's trace, we found a significant effect of interface and task difficulty on total actions and total time, $V = .967$, $F(2, 55) = 800.27$, $p < .001$, $\eta 2 = .967$. Subsequent separate univariate 3x3 ANOVAs on the dependent variables revealed more detailed results. For the case of total actions, there was a significant effect of interface and difficulty $F(8, 56) = 19.99$, $p < .001$, $\eta 2 = .741$, as well as a significant interaction effect of both variables $F(4, 56) = 3.199$, $p = .02$, $\eta 2 = .186$. Figure 2 clearly shows this interaction effect. Then, for total time (see Figure 3), we also found a significant effect of interface and difficulty $F(8, 56) = 13.78$, $p < .001$, $\eta 2 = .663$, but there was no interaction effect, $F(4, 56) = .551$, $p = .699$, $\eta 2 = .038$. Moreover, we found significant marginal effects of interface and difficulty.

For total time we tested multiples pairwise differences across interfaces and across tasks, using Bonferroni correction. Post-hoc analyses show that there is a significant difference between the total time spent on the baseline interface (M = 413.48, SE = 38.81) compared to the tag cloud interface (M = 312.4, SE = 30.74), $p = .024$. In terms of task difficulty, the result was as expected, since time spent on difficult tasks was significantly larger than medium tasks, $p < .001$, and these had larger timespan than low difficulty tasks, $p < .001$, respectively.

In summary, we need to consider two main variables to answer RQ1. Our analysis on two usual performance metrics, time to complete task and number of actions, shows that both *interface* and *task difficulty* present an interesting interplay to explain the results. On one hand, people spend more time on the baseline interface than on the tag enabled ones, particularly compared to the simpler tag interface. On the other hand, the number of actions is dependent upon task difficulty.

4.2. Log-based Pattern Analysis

To further understand the observed differences in performance, we conducted a deeper analysis of user action logs collected from the first study with 24 participants. Although the performance analysis reveals key differences between interfaces and tasks with difficulty as a contextual factor, it does not show how these general differences are composed from different kinds of user actions. To further investigate why those differences might take place, we performed a log analysis[1] of users' actions for all interfaces and tasks. While some actions are available in the three interfaces, others are only possible in the tag-enabled interfaces. The statistics on actions Search, Show more results, Click image, and Total actions can be compared across the three interfaces. The remaining actions are only present in the tag-enabled interfaces.

Considering the impact of subtask difficulty shown in the previous subsection, we conducted a 3x3 MANOVA on outcome variables search, show more results, click image and total actions as a function of the interface and task difficulty. Pillai's trace

[1] We logged users' actions during the search process including the query they issued, and all the other interactions they made such as "Add tag", "Remove Tag", etc. as presented in Table 5.

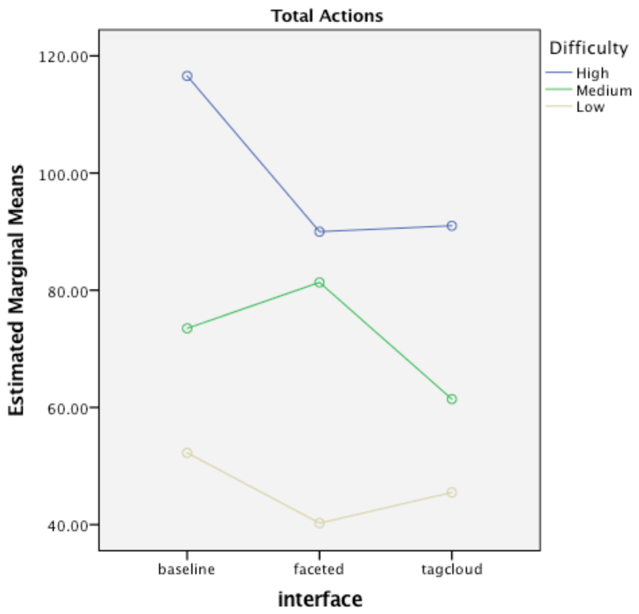

Fig. 2. Plot shows mean number of total actions per user, considering the three interfaces in the x-axis, and lines shows different level of task difficulty.

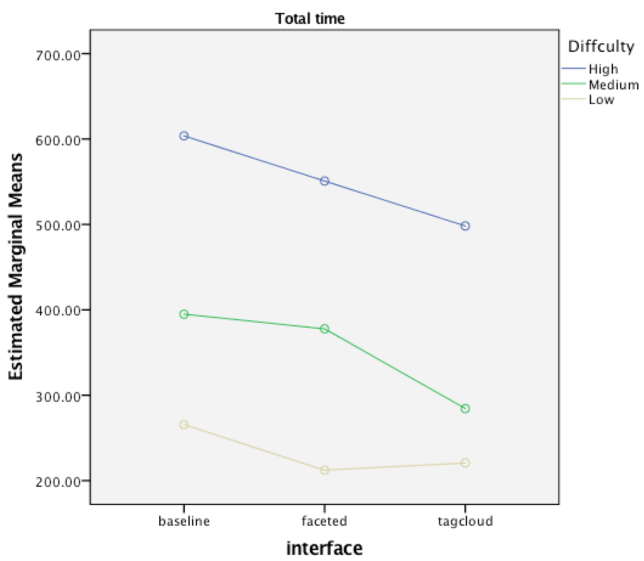

Fig. 3. Plot shows mean timespan per user (in second), considering the three interfaces in the x-axis, and lines show different levels of task difficulty.

Table 4. Mean and S.E. of different actions of the exploratory search in different subtasks (rows) at the three interfaces.

Difficulty	Measure	Baseline	Tag-cloud	Faceted tag-cloud
Low	Total time	265.63±36.06	220.75±36.06	212.25±36.06
	Search actions	9.75±1.28*	4.75±1.28	1.88±1.28
(Music)	Click tag	0	4.63±1.03	2.75±1.03
	Add tag	0	0.25±0.39	0.5±0.39
	Remove term	0	0.38±0.4	0.88±0.4
	Show more tags	0	0±0.14	0.13±0.14
	Show less tags	0	0±0.03	0±0.03
	Show more res.	0.88±0.43	0.25±0.43	0.38±0.43
	Click image	13.13±0.92	12.38±0.92	12±0.92
	Total actions	23.75±2.44	22.63±2.44	18.5±2.44
Medium	Search time	394.88±36.06	284.43±38.55	377.78±34
	Search actions	18±1.28	10.29±1.37*	15.22±1.21
(Sports)	Click tag	0	5.43±1.1	11.33±0.97***
	Add tag	0	0.29±0.42	0±0.37
	Remove term	0	0.86±0.43	2.67±0.38
	Show more tags	0	0±0.15	0.44±0.13
	Show less tags	0	0±0.03	0±0.03
	Show more res.	0.25±0.43	0±0.46	0.89±0.41
	Click image	20.5±0.92	17.57±0.98	21.89±0.87
	Total actions	38.75±2.44	34.43±2.61	52.44±2.3***
High	Search time	603.71±38.55	498.2±45.61	550.8±45.61
	Search actions	28.71±1.37***	9±1.62***^	19.4±1.62^
(Stores)	Click tag	0	12.6±1.3*	8±1.3
	Add tag	0	0.6±0.49	1.2±0.49
	Remove term	0	1.6±0.51	0.8±0.51
	Show more tags	0	0.2±0.18	0±0.18
	Show less tags	0	0±0.04	0±0.04
	Show more res.	3.29±0.46	0.8±0.55	1±0.55
	Click image	35.14±0.98***	24.8±1.16	24.8±1.16
	Total actions	67.14±2.61**	49.6±3.09	55.2±3.09

*=significant at $p < .05$; **=significant at $p < .01$; ***=significant at $p < .001$.

test shows a significant effect of interface and task difficulty on the outcome variables V = .965, F(5, 52) = 283.13, $p < .001$, $\eta 2 = .965$. We further conducted univariate ANOVAs on the outcome variables and we found that all were significant. More in detail, while variables show more results F(4, 56) = 4.39, $p = .112$, $\eta 2 = .123$ and search time F(4, 56) = 5726.84, $p = .699$, $\eta 2 = .038$ presented no interaction effect of interface and difficulty, the other IVs search actions, F(4, 56) = 129.904, $p = .003$,$\eta 2 = .249$, total actions F(4, 56) = 479.52, $p = .008$, $\eta 2 = .215$, and click image actions F(4, 56) = 80.92, $p = .02$, $\eta 2 = .185$, resulted with interaction effects. These results show again the interplay of task difficulty with tag enabled interfaces to explain the effects observed in our experiments. The detailed mean and standard errors of each condition are shown in Table 5.

This analysis confirms the results shown in the previous MANOVAs. An interesting outcome is that although click image actions did not show differences between interfaces, it become evident when introducing the task difficulty as in Figure 5. On tasks with high difficulty the users of the baseline interface click consistently more on images than users of the tag enabled interfaces. This result points to an important role of tags, which help not only in retrieving, but also in understanding the results. Without additional support provided by tags, participants have a stronger need to access the image details to examine relevance and find new ideas for query terms, while in the presence of tags that provide necessary information without image access this need is reduced.

Altogether, time and click-log analysis registered important performance and usage pattern differences between the three explored interfaces. The data pointed to some possible reasons of the observed efficiency of tag-enabled interfaces and provided some evidence that the value of tags go beyond usage of tags. To explain some differences observed in a first study, we decided to perform the second study using eye-tracking analysis as the primary data collection tool. Other important points to consider as results:

a) These results point to a first reason of higher efficiency of tag-enabled interfaces: tag clicking is usually faster than formulating or typing a query.
b) By categorizing the difference in the number of actions by type, we show that the most significant reduction between search-based and faceted interfaces is observed in the number of search actions. This points to another reason for higher efficiency of tag-enabled interfaces: the presence of tags made users' search attempts not only faster, but also more efficient, either by providing alternative ideas for query formulations or by offering meaningful terms to click. Constructing a good query is hard; the presence of tags makes it easier, significantly decreasing the fraction of unsuccessful search cases when the users were not able to get relevant results on the top of the results page and asked for more.
c) In addition, the lower rate of reduction of search actions, image clicks and requests for more results for the case of faceted interface, provides some evidence that the faceted tag-cloud interface was a bit more difficult or confusing for the users in the search process. It was also causing more needs to "undo" their previous actions. The analysis of the most visible difference between the tag-enabled interfaces (for total retrieval attempts) confirms this hypothesis (Figure 4). For the majority of users, the effectiveness of the faceted interface was comparable with the effectiveness of the tag cloud. However, for some fraction of users, the faceted interface required considerably more effort to achieve the goal (seven users with the faceted interface against just one with the tag cloud interface making over 25 retrieval attempts). This data leads to the hypothesis that the faceted interface might be specifically problematic for some users, for one of the tasks, or a combination of these factors. To examine this hypothesis, we compared user performance and actions between tasks.

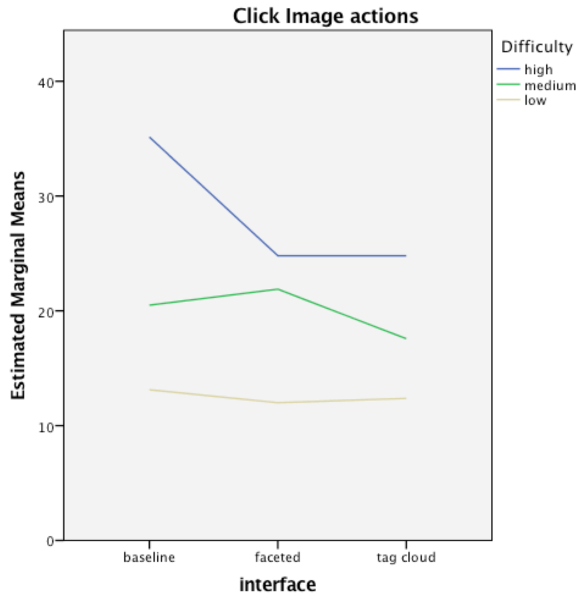

Fig 4 Plot show the trend of mean "click image" actions per user considering different interface (x-axis) and difficulty level (lines).

4.3 Eye-tracking Pattern Analysis

The eye-tracking analysis is based on the data of nine participants collected in our second study. The participants performed the same experiment procedure while being observed by a Tobii 1750 eye tracker. As mentioned above, the interfaces were slightly modified to prevent the participants from scrolling the pages up or down. It was necessary to conduct the posterior analysis of eye-tracking data, since the Tobii software tracked eye fixations relative to the upper-left

- Area of Interest (AOI): A well-defined region in the interface where the user looked at while performing a task.
- Fixation: A moment where the eye is relatively motionless [60] and is taking in or encoding information [61].
- Saccade: Eye movements occurred between fixations [61].
- Transition: A saccade between two AOIs.

We defined seven AOIs (see Table 6). As explained above, each version of the search interfaces included two pages: results (the page where the thumbnails of images found in response to a query are listed) and detail (where a single image with its details is shown). These AOIs are shown in the context of the interfaces in Figure 6 and 7 identifying them with bounding boxes. The common AOIs (implemented in all three interfaces) are marked in blue, while tag-specific AOIs are marked in purple.

Table 5. List of AOIs with the corresponding letter used.

Letter	Area of Interest	Page	Interface
A	Task Description	Results, Detail	Baseline, Tag-cloud, Faceted
Q	Query & Query expansion	Results, Detail	Baseline, Tag-cloud, Faceted
R	Results & Number of results	Results	Baseline, Tag-cloud, Faceted
C	Collection Box (where the user places images found)	Detail	Baseline, Tag-cloud, Faceted
D	Image Detail	Results, Detail	Baseline, Tag-cloud, Faceted
T	Tag cloud or faceted tag clouds (who, when, what, where, other)	Results, Detail	Tag-cloud (a single AOI), Faceted (5 sub-AOIs combined)
H	Help	Results, Detail	Baseline, Tag-cloud, Faceted

Fixation analysis

Two features, the number of fixations and fixation duration, are commonly used in eye tracking analysis [62]. As outlined in [61], we assume that higher fixation frequency and longer fixation duration indicate less efficient search or more difficulty in extracting information with the particular interface. In this subsection, we present our analysis of eye-tracking data for the exploratory search task only. While we performed the analysis for the lookup search as well, no significant differences or interesting patterns were found for this type of task. We first conducted a one-way ANOVA on both the fixation duration and the number of fixations among three interfaces, but we did not find significant differences for the fixation duration, $F(2, 24) = .279$, $p = .759$, and the number of fixations, $F(2, 24) = .191$, $p = .828$. Therefore, we continued a more in-depth analysis by splitting the dataset to different tasks.

To explore whether subjects performed significantly different in three interfaces with three difficulty tasks, we performed 3x3 between-subjects ANOVA on the number of fixations (there are three participants in each condition). Significant differences were found on the number of fixations among interfaces for the most complicated task (stores), $F(2, 18) = 7.139$, $p = .005$, partial $\eta^2 = .442$ (Table 7). Users working with the stores task using the tag-cloud interface (M= 680.33, SE = 173.632) had a significantly lower number of fixations than the users working with the same task using the baseline (M = 1041.00, SE = 181.045), $F(1, 18) = 5.781$, $p = .027$, partial $\eta^2 = .243$, as well as a significantly lower number of fixations than the users working with the faceted tag-cloud (M = 1239.33, SE = 106.068), $F(1, 18) = 13.887$, $p = .002$, partial $\eta^2 = .436$. There were no significant differences on the number of fixations among interfaces for simple and average complexity tasks sports and music. This suggests that users obtained more support from the tag-cloud interface than they received from either the baseline or the faceted tag-cloud; hence, they need to allocate less attention with the tag-cloud interface.

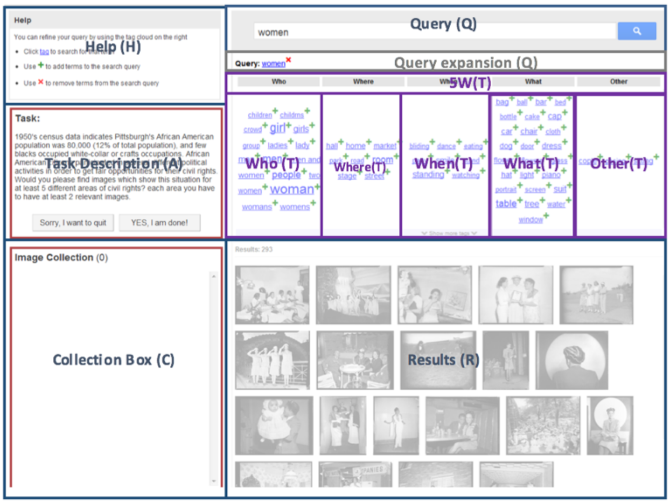

Fig. 5. Area of Interests (AOI letter) in the faceted tag-cloud interface (Result page)

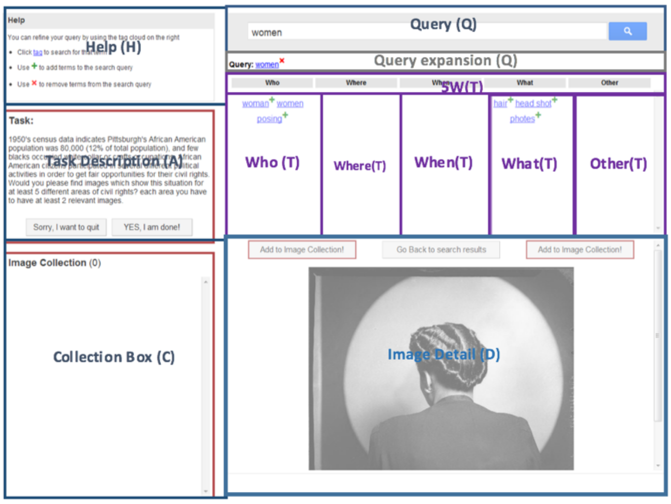

Fig. 6. Area of Interests (AOI letter) in the faceted tag-cloud interface (Detail page)

Table 6. Descriptive statistics (mean±SE) of number of fixations and fixation duration by task and interface.

Difficulty	Measure	Baseline	Tag-cloud	Faceted tag-cloud
Low (Music)	number of fixations	333.00±141.23	478.67±20.38	403.00±64.73
	Fixation duration	94.45±44.08	111.95±13.117	105.91±18.28
Medium (Sports)	number of fixations	611.67±68.69	748.67±43.22	548.33±79.72
	Fixation duration	168.37±35.10	188.72±21.93	143.71±25.39
High (Stores)	number of fixations	1041.00±181.05*	680±173.63*^	1239.33±28.88^
	Fixation duration	253.89±36.35	201.95±47.30	342.60±2.61

* ,^=significant differences.

Table 7: The distribution of AOIs fixations in the baseline, tag-cloud, and faceted tag-cloud interfaces

AOI	Baseline	Tag-cloud	Faceted tag-cloud
Image_detail	1914	1547	1672
Query	888	451	346
Result	1650	1464	1704
Number_of_results	108	33	41
Collection_box	498	504	677
Task_description	372	426	546
Query_expansion	0	201	119
Tag_cloud	0	738	1032

In terms of fixation duration, there was no difference among interfaces and tasks. The main effect of tasks was detected $F(2, 18) = 21.53$, $p < .001$, partial $\eta^2 = .705$. No other effect was found. Post-hoc analysis on the effect of tasks showed that the most difficult task, "stores", had significantly longer fixation duration than both other tasks. After checking the effects of interfaces on the number of fixations, we examined the effects of each AOI in different interfaces in order to get more insights of the interface elements. According to [63], more fixations on a particular area indicate that it is more important, visible, and interesting to the viewer than other areas. Table 8 shows the distribution of the fixations by AOIs. We summed the number of fixations of each facet-related AOI (who, where, when, what, other, and 5w) in the faceted tag-cloud interface to compare with the single Tag_cloud AOI in the tag-cloud interface.

The most important message provided by this table is that a very large number of fixations for tag-enabled interfaces was made in the tag cloud area, lagging behind only the key examination area on each page (details and result), but overshadowing such important areas as Task_description, Collection_box, and Query. Specifically, comparing the fixation on Query and tag areas with the number of retrieval actions starting from these areas shown in Table 3, we can observe that the number of fixations in the tag areas is disproportionally high. While the number of tag clicks in both tag-enabled interfaces is considerably smaller than the number of queries, the number of fixations in the tag area is much larger than in the query area. This is a clear indication that tags were more than just a place to find a term to click. This supports our earlier hypotheses that tags helped with the analysis of search results and formulating better queries. More detailed analysis of the role of tags is provided in the next section. In addition, comparing the number of tag clicks in both tag-enabled interfaces, the increase in the number of fixations from tag cloud to faceted tag-cloud is disproportionally high. While the number of clicks is just a bit higher, the number of fixations increased 50%.

To examine the difference in AOI fixations between tasks and interfaces, we conducted 2x3 between-subjects ANOVA on the number of fixations on each tag-related AOI. We found a significant difference only for the Tag_cloud AOI based on the main effect of task, $F(2, 18) = 6.607$, $p < .012$, partial $\eta^2 = .524$. The faceted tag-cloud interface ($M = 644$) attracted a significantly higher number of fixations in the Tag_cloud AOI than the tag-cloud interface ($M = 237$) for task "stores," $F(2, 18) = 10.032$, $p < .008$, partial $\eta^2 = .455$. As above, we speculate that the faceted tag-cloud offer more functionalities in exploratory search tasks causing the users to refer to it more frequently. The result hints that this functionality might be especially important to users when they perform more complicated tasks. At the same time, it might be evidence that the simple tag-cloud interface was easier for users to operate without additional distraction caused by the many facets provided in the faceted tag-cloud interface.

While the analysis of fixations hints that tags play several important roles and that the faceted tag-cloud might provide broader support, fixations alone cannot provide sufficient evidence. To understand how the tags were used in the process of exploratory search, we need to perform the transition analysis of the eye-tracking data.

Transition analysis

The eye-tracker software captured a complete sequence of fixations for each task performed by each participant. The transition analysis examines each participant's eye movement from one AOI to another. The frequency and probability info of transition between AOIs are captured by ClearView for the further data analysis. In order to keep the readability of the analysis of gaze transitions, we only focused on the transition pairs among Query (Q), Result(R), Image_detail (D), and Tag_cloud (T) AOIs. The remaining AOIs were merged into one state marked others (O) to show transitions from QRDT to the other states and transitions from other states were ignored. Figure 8 shows two state transition diagrams - one for the result page and another for the detail page of the two tag-based interfaces. The state diagrams display the relative transition frequencies (in percentages), i.e., the frequency of each sum of the frequencies of QT, QR, QO, RT, RQ, RO, TQ, TR, TO).

- The dominated eye-movement transition for the *detail page* in both interfaces is DT, which indicates users attempt to associate the detailed image with tags that describe it. Arguably, the faceted interface that categorizes tags provides slightly better support for this task since the users refer to it more frequently (19% and 20% vs. 15% and 17%). In contrast, users of the tag-cloud interface refer more frequently to the query in their attempts to interpret the image (9% and 11% vs. 5% and 6%).
- The dominated eye-movement transition in the result page of tag-cloud interface is QT. Users move their eyes between Q and T more frequently than between other AOI and between the same AOIs in the tag-cloud interface. This might indicate that the regular tag cloud is more useful in re-formulating queries, possibly serving as a source promising query terms.
- The dominated eye movement transition for the result page of the faceted tag-cloud interface is RT. Users move their eyes between R and T more frequently in this interface than between other AOIs as well as between R and T in the tag-cloud interface. Given that RT transition indicates user attempts to interpret search results in terms of tags, this data provides another evidence that faceted tag-cloud interface provides better support in interpreting and understanding results than the regular tag cloud interface.

To better understand difference in patterns of user behavior we also produced separate transition diagrams for top and bottom performers based on their search time (Figure 9). This was important since log analysis indicated considerable performance differences between users, especially for the faceted interface (Figure 4). As top performers for each interface we selected three users (out of nine) who spent the least

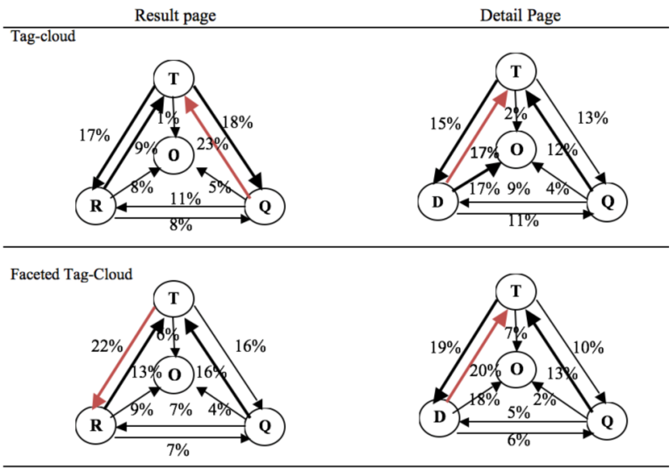

Fig. 7. State diagrams of gaze transitions among AOIs on the result page (left) and detail page (right) in two tag-based interfaces. The numbers along each arrow shows the percentage of the transition frequency. The arrow marked with red color is the most popular transition.

time to finish the exploratory task using this interface. As bottom performers for each interface we selected three users who spent the most time to finish it with the interface. The analysis of behavior of top and bottom performers provides additional support for the insights provided by the previous analysis. As we can see, interpreting search results in terms of tags (RT transition on the results page, Figure 11) was a successful strategy for the faceted tag-cloud interface, which, arguably, provided good support for this task. In contrast, a more successful strategy for the tag-cloud interface was shifting attention to matching results with the query (RQ transition on the results page, Figure 10). The use of RT transition at the expense of RQ transition in a regular tag-cloud interface was associated with low performance - further evidence that the regular tag-cloud interface provides less support for understanding and interpreting search results.

On the other hand, QT transition was considerably important for both kinds of interfaces and both groups of participants. The focus of this transition was to understand the "aboutness" of the query results in terms of tags (QT) and to spot helpful terms for the future queries (TQ). As argued earlier, the higher frequency of using this transition in the tag cloud interface might indicate that these processes were better supported by the simpler tag cloud interface. A more detailed analysis provided by the split diagrams indicates that the critical factor here was better support for tag reuse and query reformulation and the more efficient users of the tag cloud interface were able to better use this advantage (largest fraction of TQ transition). In contrast, the increase of QT transfers was connected to poor performance, which is another indication that tag cloud is less helpful in comprehending results.

To further explore whether tag cloud is generally more effective for tag-reuse in query reformulation and top performers could better use this support to their advantage, we examined the queries issued by the top and bottom performers in both tag-enabled interfaces and if any visible tags in their queries. We checked the overlap between queries and tags by calculating the fraction of queries that contained terms displayed in

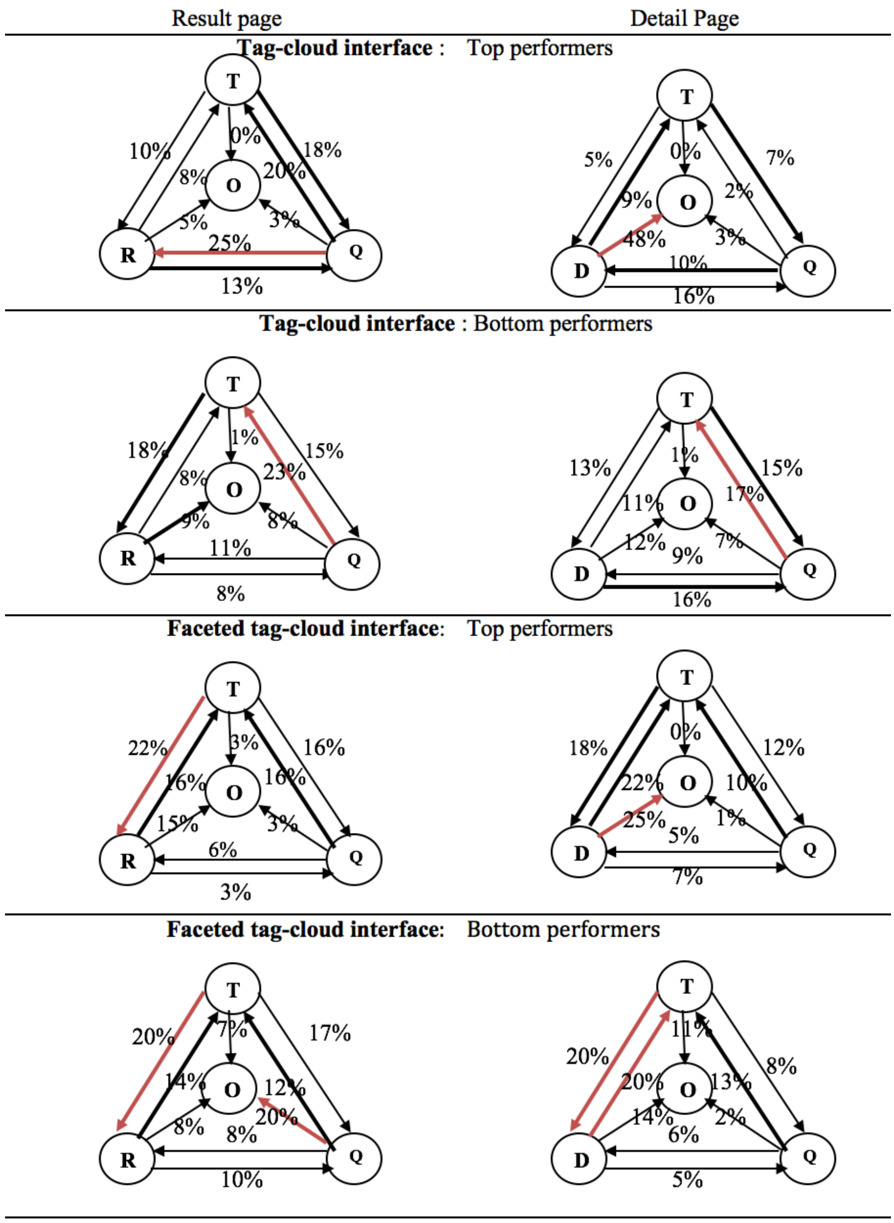

Fig. 8. State diagrams of gaze transitions of top performers and bottom performers among AOIs on the result page and detail page in two tag-based interfaces. The numbers along each arrow shows the percentage of the transition frequency.

the tag cloud among all issued queries. Figure 9 shows the results presenting the ratio by all participants, top and bottom performers, in both interfaces. One t-test's results indicated that the overlap between search terms and tags was marginally higher in the

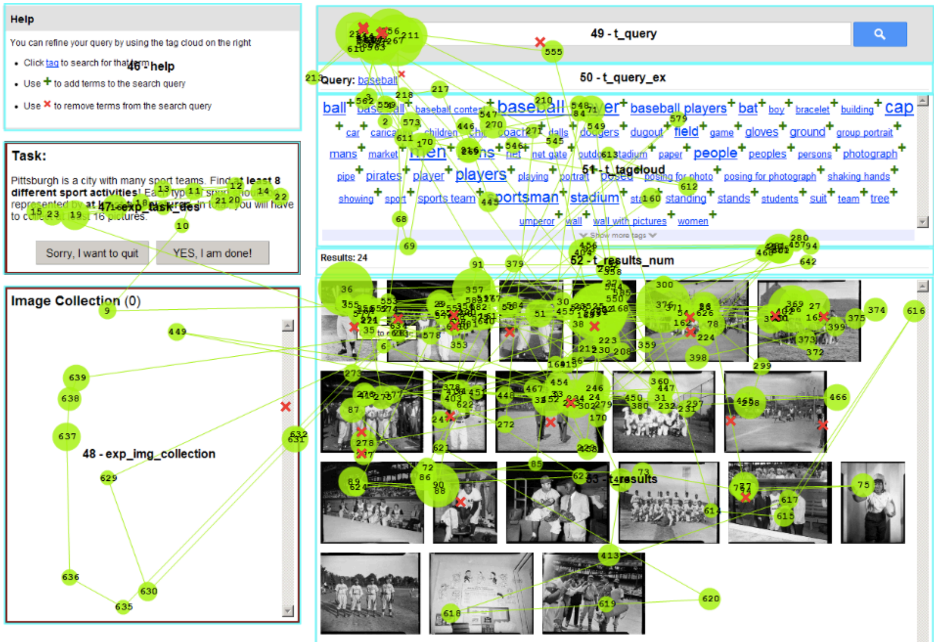

Fig. 9. Screenshot with scanpath between Query (Q) and Results (R) AOIs of a top performer in the tag-cloud interface.

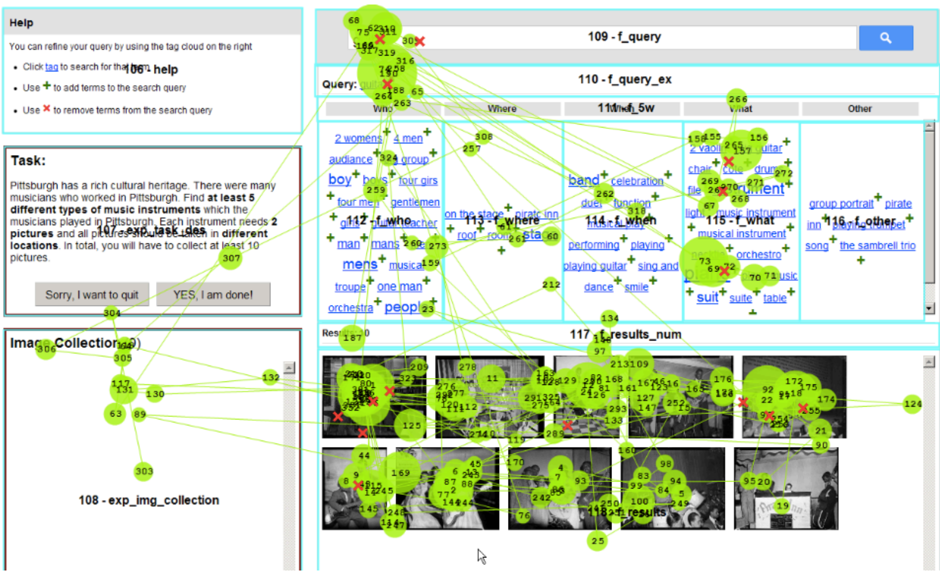

Fig. 10. Screenshot with scanpaths highlighting the Query-Tag (QT) pattern (red) and Tag-Results (TR) pattern (blue) of a top performer in the faceted tag-cloud interface.

tag cloud than in the faceted tag-cloud, p = .7. Another t-test comparing the top and bottom performers show that the overlap was significantly higher among top performers than among bottom performers in the tag-cloud interface, p = .03. This confirms that tag cloud provided better support for tag reuse and that top performers in that interface were able to leverage this support in their favor. The hypothesis that the regular tag-cloud interface provides better support for search and query formulation while the faceted tag-cloud interface provides better support for understanding and interpreting search results could explain several results reported above. For example, it explains the larger number of fixations in the tag area for the most difficult exploratory search where understanding and interpreting results was most critical to success. It also helps to explain why the tag-cloud interface with better search support was most efficient overall, while the faceted tag-cloud delivered a slightly better success rate.

5. Conclusion and Discussion

The study presented focused on the value of tags in interfaces for accessing image collections. We attempted a broad exploration comparing traditional search-only access with regular and faceted tag-cloud in lookup and exploratory search.

5.1 Performance differences between the interfaces

The bottom level performance data collected in the study demonstrated that tag-enabled interfaces are significantly more efficient in more complex exploratory search tasks. The users of both tag-enabled interfaces were able to complete exploratory tasks using significantly fewer actions and, in case of a tag-cloud interface, significantly faster. No significant performance differences were found between all three interfaces for lookup search demonstrating that the advantage provided by tags is critical for more challenging situations.

While the faceted interface was less efficient on average than the traditional tag-cloud in the exploratory search context, the performance differences between the two tag-enabled interfaces were not significant. We also observed a broader distribution of performance parameters for the faceted one, indicating that it was considerably more challenging for some users or tasks. These effects were explained when we separately compared the performance of three interfaces on the tasks with significantly different levels of difficulty. The comparison showed that the traditional tag-cloud performs reliably better than the search-only interface increasing its advantage with the increase of task difficulty and reaching significance for the most complex task. In contrast, the advantage of tags in the faceted tag-cloud interface was not stable. On the easiest music task, it was on average most efficient; however, on the medium difficulty sport task, while the average is still faster than search-only, it required significantly more actions than both search-only and the regular tag-cloud interface. This data hints that the efficiency of the faceted tag-cloud interface does depend on the nature of the task. For some tasks, the faceted organization provide a performance boost, while for others the advantage could be minimal since the complex organization could add confusion.

5.2. Specific patterns for the differences between the interfaces

The log-based pattern analysis provided good evidence that one reason of higher efficiency of tag-enabled interfaces stems from the ability to combine queries and tag clicks for image retrieval. While the total number of retrieval attempts (queries plus tag clicks) was comparable in the three interfaces, the users of both tag-enabled interfaces performed significantly fewer searches replacing a considerable number of searches with tag clicks. Link clicking is known as more efficient since the days of hypertext research not just because of slower query typing, but also due to recognition vs. recall cognitive difference. While tag clicks are based on faster memory recall, query formulation requires slower term retrieval from memory. Query formulation is known to be quite challenging especially for complex exploratory search tasks, which helps to understand why significant performance differences were found for more complex tasks.

At the same time, the log-based pattern analysis demonstrated that it is not just faster clicks that made tag-enabled interfaces more efficient. We found that the users of both tag-enabled interfaces were able to complete their search tasks while examining significantly fewer images. The action-level analysis can't answer why the users might need to examine images – it might be to analyze search results, to judge image relevance, or to find good tags for future queries. However, a significant decrease of these actions means that these needs were better supported in the presence of tags.

The eye-tracking pattern analysis provided further insights on the role of tags in exploratory search. Fixation analysis indicated that tag areas attracted a lot of user attention, being the second most important area in each of the two analyzed screens after the results component. We also observed that the number of fixations in the tag areas was much higher than expected by the analysis of tag clicking actions in these areas. Indeed, the number of tag clicks in both tag-enabled interfaces was considerably smaller than number of queries, while the number of fixations in the tag area was much larger than in the query area. These data provided further evidence that tags performed several functions in the exploratory search process, beyond their original function as click points. The fixation analysis also indicated that the faceted tag-cloud attracted more fixations than the traditional tag-cloud. This difference reached significance for the most complex stores task. This indicated that the interface might support broader set of needs while being also more challenging and confusing to operate.

To better understand how tags were used in the process of exploratory search in two tag-enabled interfaces, we complemented basic fixation analysis with transition analysis. The transition analysis uncovered several eye-movement patterns that, as we believe, indicated that tags were used for reflecting on a query in terms of produced tag cloud, interpreting search results (both list of images and individual images), and as a source of ideas for new query formulation. Moreover, the transition frequency analysis along with query formulation analysis indicated that the traditional tag-cloud interface provided significantly better support for query formulation while the faceted tag-cloud interface provided better support for interpreting search results. We believe that this hypothesis fits well with several observed differences between traditional and faceted tag-clouds.

By combining several kinds of data analysis, our study demonstrated the added value of the tag-enabled interfaces in the context of more complex exploratory search tasks and uncovered several reasons for the observed performance difference. It also highlighted differences between two explored tag-enabled interfaces and connected them with interface usage patterns and user feedback.

5.3. Limitations

Although there are still several limitations in our studies, neither the collection we used, nor the tasks were sufficiently favorable for the faceted organization of tags limiting our ability to determine its true impact. Since the collection had no "native" tags provided by real users, the crowdsourcing approach by recruiting Mechanical Turkers has been examined in [47] showing no effect on image search performance. While some findings have been confirmed by significance analysis, some others should be still considered as hypotheses and will need additional studies to be confirmed. We hope to address these issues in our future work.

References

[1] A. J. Munro, K. Höök, D. Benyon, and (Eds.), *Social navigation of information space*. Springer Science & Business Media., 2012.

[2] Y. Ding, E. K. Jacob, E. Yan, N. L. George, and L. Guo, "Perspectives on social tagging," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 12, pp. 2388–2401, 2009.

[3] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve Web search?," in *Proceedings of the international conference on Web search and web data mining*, 2008, pp. 195–206.

[4] J. Sinclair and M. Cardew-Hall, "The folksonomy tag cloud: When is it useful?," *J. Inf. Sci.*, vol. 34, no. 1, pp. 15–29, May 2008.

[5] D. Skoutas and M. Alrifai, "Tag clouds revisited," in *Proceedings of the 20th ACMInternational Conference on Information and Knowledge Management, CIKM2011*, 2011, pp. 221–230.

[6] B. Lee, N. H. Riche, A. K. Karlson, and S. Carpendale, "Sparkclouds: Visualizing trends in tag clouds," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1182–1189, 2010.

[7] C. Collins, F. B. Viegas, and M. Wattenberg, "Parallel tag clouds to explore and analyze faceted text corpora," in *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on.*, 2009, pp. 91–98.

[8] G. Koutrika, Z. M. Zadeh, and H. Garcia-Molina, "Data clouds: summarizing keyword search results over structured data," *Proc. 12th Int. Conf. Extending Database Technol. Adv. Database Technol.*, pp. 391–402, 2009.

[9] M. Halvey and M. T. Keane, "An assessment of tag presentation techniques," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 1313–1314.

[10] 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 ACM Conference on Human Factors in Computing Systems*, 2007, pp. 995–998.

[11] 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*, 2011, pp. 835–844.
[12] S. Lohmann, J. Ziegler, and L. Tetzlaff., "Comparison oftag cloud layouts: Task-related performance and visual exploration.," in *IFIP Conference on Human-Computer Interaction.*, 2009, pp. 392–404.
[13] S. Bateman, C. Gutwin, and M. Nacenta, "Seeing things in the clouds : The effect of visual features on tag cloud selections," in *the 9th ACM conference on Hypertext and hypermedia*, 2008, pp. 193–202.
[14] J. Bar-Ilan, M. Zhitomirsky-Geffet, and Y. Miller, "Tag-based retrieval of images through different interfaces: A user study," *Online Inf. Rev.*, vol. 36, no. 5, pp. 739–757, 2012.
[15] B. Y. Kuo, T. Hentrich, B. M. Good, and M. D. Wilkinson, "Tag clouds for summarizing Web search results," in *Proceedings of the 16th international conference on World Wide Web (WWW '07)*, 2007, pp. 1203–1204.
[16] D. Wu, D. He, J. Qiu, R. Lin, and Y. Liu, "Comparing social tags with subject headings on annotating books: A study comparing the information science domain in English and Chinese," *J. Inf. Sci.*, vol. 39, no. 2, pp. 169–187, Oct. 2012.
[17] D. Dimitrov, D. Helic, and M. Strohmaier, "Tag-based navigation and visualization," in *Social Information Access*, Springer, Cham., 2018, p. 181–212.
[18] M. Gupta, R. Li, Z. Yin, and J. Han, "An overview of social tagging and applications," in *Social Network Data Analytics*, C. C. Aggarwal, Ed. Boston, MA: Springer US, 2011, pp. 447–497.
[19] B. N. Bullock, A. Hotho, and G. Stumme, "Accessing Information with Tags: Search and Ranking," in *Social Information Access*, Springer, Cham., 2018, pp. 310–343.
[20] P. Brusilovsky and D. He, *Social Information Access: Systems and Technologies.*, vol. 10100. Springer, 2018.
[21] G. W. Furnas *et al.*, "Why do tagging systems work?," *CHI '06 extended abstracts on Human factors in computing systems - CHI '06*. ACM Press, New York, New York, USA, p. 36, 21-Apr-2006.
[22] O. Nov and C. Ye, "Why do people tag? Motivations for photo tagging," *Commun. ACM*, vol. 53, no. 7, pp. 128–131, Jul. 2010.
[23] M. Strohmaier, C. Körner, and R. Kern, "Why do users tag? Detecting users' motivation for tagging in social tagging systems," in *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2010.
[24] Y. Hassan-Montero and V. Herrero-Solana, "Improving tag-clouds as visual information retrieval interfaces," in *International Conference on Multidisciplinary Information Sciences and Technologies*, 2006, pp. 25–28.
[25] 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*, 2011, p. 1.
[26] 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 andData Mining, WSDM 2011*, 2011, pp. 835–844.
[27] M. Lipczak and E. Milios, "Efficient tag recommendation for real-life data," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 1, pp. 1–21, Oct. 2011.
[28] K. Dellschaft and S. Staab, "Measuring the Influence of Tag Recommenders on the Indexing Quality in Tagging Systems.," in *Proceedings of the 23rd ACM conference on Hypertext and social media*, 2012, pp. 73–82.
[29] Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang., "Personalized tag recommendation using graph-based ranking on multi-type interrelated objects," *SIGIR'09 Proceedings of the 32nd annual international ACM SIGIR conference on research and development in information retrieval*. pp. 540–547, 2009.

[30] D. Helic, C. Körner, M. Granitzer, M. Strohmaier, and C. Trattner, "Navigational efficiency of broad vs. narrow folksonomies.," in *Proceedings of the 23rd ACM conference on Hypertext and social media*, 2012, pp. 63–72.
[31] M. Leginus, P. Dolog, and R. Lage, "Graph based techniques for tag cloud generation," in *Proceedings of the 24th ACM conference on hypertext and social media*, 2013, pp. 148–157.
[32] M. Leginus, C. Zhai, and P. Dolog, "Personalized generation of word clouds from tweets," *J. Assoc. Inf. Sci. Technol.*, 2015.
[33] C. Trattner, D. Helic, and M. Strohmaier, *Tag Clouds*. Springer New York, 2014.
[34] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins, "Visualizing tags over time.," *ACM Trans. Web*, vol. 1, no. 2, p. 7, 2007.
[35] T. Russell, "Cloudalicious: Folksonomy over time.," in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2006*, 2006, p. 364.
[36] E. H. Chi and T. Mytkowicz, "Understanding navigability of social tagging systems," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2007, vol. 7.
[37] M. Hearst and D. Rosner, "Tag clouds: Data analysis tool or social signaller?," in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, 2008, pp. 160–169.
[38] C. Trattner, Y. Lin, D. Parra, and P. Brusilovsky, "Evaluating tag-based information access in image collections," in *Proceedings of the 23rd ACM conference on Hypertext and social media*, 2012, pp. 113–122.
[39] G. Smith, *Tagging: People-powered metadata for the social Web*. New Riders, 2007.
[40] J. Schrammel, M. Leitner, and M. Tscheligi, "Semantically structured tag clouds : An empirical evaluation of clustered presentation approaches," in *International Conference on Human Factors in Computing Systems*, 2009, pp. 2037–20740.
[41] E. R. Tufte, "Beautiful Evidence," in *Graphics Press*, 2006.
[42] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu, "Can all tags be used for search ?," in *Proccedings of the CIKM conference of Information and Knowledge Management*, 2008, pp. 203–212.
[43] D. Böhnstedt, L. Lehmann, C. Rensing, and R. Steinmetz, "Automatic identification of tag types in a resource-based learning scenario," in *the 6th European Conference on Technology Enhanced Laearing*, 2011, no. September, pp. 57–70.
[44] M. Hearst, "Clustering versus faceted categories for information exploration," *Commun. ACM*, vol. 49, no. 4, p. 59, Apr. 2006.
[45] R. Capra and G. Marchionini, "The relation browser tool for faceted exploratory search," in *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, 2008, pp. 420–420.
[46] B. Sigurbjörnsson and R. van Zwol, "Tagexplorer : Faceted browsing of flickr photos," 2010.
[47] Y.-L. Lin, P. Brusilovsky, and D. He, "Finding cultural heritage images through a Dual-Perspective Navigation Framework," *Inf. Process. Manag.*, vol. 52, no. 5, pp. 820–839, 2016.
[48] C. di Sciascio, P. Brusilovsky, and E. Veas, "A study on user-controllable social exploratory search," in *23rd International Conference on Intelligent User Interfaces*, 2018, pp. 353–364.
[49] Y. L. Lin, C. Trattner, P. Brusilovsky, and D. He, "The impact of image descriptions on user tagging behavior: A study of the nature and functionality of crowdsourced tags," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 9, pp. 1785–1798, 2015.
[50] K. Byström and K. Järvelin, "Task complexity affects information seeking and use," *Inf. Process. Manag.*, vol. 31, no. 2, pp. 191–213, 1995.
[51] A. Diriyey, A. Blandfordy, and A. Tombrosz, "When is system support effective?," in

the third symposium on Information interaction in context, 2010, pp. 55–64.

[52] J. Jiang, D. He, and J. Allan, "Searching, browsing, and clicking in a search session: Changes in user behavior by task and over time," in *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2014, pp. 607–616.

[53] M. Hearst, *Search user interfaces*. Cambridge University Press, 2009.

[54] G. Marchionini, "Exploratory search: From finding to understanding," *Commun. ACM*, vol. 49, no. 4, pp. 41–46, Apr. 2006.

[55] A. Gegenfurtner, E. Lehtinen, and R. Säljö, "Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains," *Educ. Psychol. Rev.*, vol. 23, no. 4, pp. 523–552, Jul. 2011.

[56] Y. Lin, J.-W. Ahn, P. Brusilovsky, D. He, and W. Real, "Imagesieve: Exploratory search of museum archives with named entity-based faceted browsing," in *Proceedings of the American Society for Information Science and Technology*, 2010, vol. 47, no. 1, pp. 1–10.

[57] J.-W. Ahn, P. Brusilovsky, J. Grady, D. He, and R. Florian, "Semantic snnotation based exploratory search for information analysts," *Inf. Process. Manag.*, vol. 46, no. 4, pp. 383–402, 2010.

[58] J. Ahn, P. Brusilovsky, and S. Han, "Personalized search: Reconsidering the value of open user models," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 2015, pp. 202–212.

[59] B. Steichen, G. Carenini, and C. Conati, "User-adaptive information visualization - Using eye gaze data to infer visualization tasks and user cognitive abilities," in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, 2013, pp. 317–328.

[60] C. Ehmke and S. Wilson, "Identifying Web usability problems from eye-tracking data," in *Proceedings of the 21st British HCI Group Annual Conference on People and Computers*, 2007, pp. 119–128.

[61] A. Poole and L. J. Ball, "Eye tracking in human-computer interaction and usability research: current status and future prospects," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, H. D. Ralph Radach, Jukka Hyona, Ed. Amsterdam: Elsevier Science, 2010, pp. 573–605.

[62] J. H. Goldberg, M. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky, "Eye tracking in Web search tasks : Design implications," in *Proceedings of the 2002 symposium on Eye tracking research & applications*, 2002, no. 650, pp. 51–58.

[63] A. Poole, L. J. Ball, and P. Phillips, "Eye-movement analysis of bookmark recognition," in *People and Computers XVIII - Desgin for Life: Proceedings of HCI 2004*, R. R. Sally Fincher, Panos Markpoulos, David Moore, Ed. London: Springer-Verlag Ltd, 2004.