An Exploratory Study of Tag-based Visual Interfaces for Searching Folksonomies

Javier Diaz  
Department of Computer Science  
University of Victoria  
PO Box 3055, STN CSC  
Victoria, BC, Canada V8W 3P6  
+1 (250) 472-5749  
diazruva@cs.uvic.ca

Keyun Hu  
Department of Computer Science  
University of Victoria  
PO Box 3055, STN CSC  
Victoria, BC, Canada V8W 3P6  
+1 (250) 472-5747  
hukeyun@cs.uvic.ca

Melanie Tory  
Department of Computer Science  
University of Victoria  
PO Box 3055, STN CSC  
Victoria, BC, Canada V8W 3P6  
+1 (250) 472-5798  
mtory@cs.uvic.ca

ABSTRACT
Aesthetic features such as animation, 3D interaction, and visual metaphors are becoming commonplace in multimedia search interfaces. However, it is unclear which attributes are needed to encourage people to use these interfaces on an ongoing basis. To design a visual interface that will elicit continual use, we first need to establish a better understanding of users’ goals and strategies, in order to determine which features are critical to support those tasks. This paper reports on an exploratory study of individuals engaging with five different image and video search interfaces. Our study helped us to understand users’ experiences with a variety of features and design elements, as well as categorize their common search tasks and strategies. We identified four distinct types of search: Search Known Objects + Known Keywords, Search Known Objects + Unknown Keywords, Search Unknown Objects + Known Keywords, and Search Unknown Objects + Unknown Keywords. We also identified common strategies used to accomplish each of these search types. Our findings suggest that search interfaces should maximize screen space used for visual representations of the media, provide on-demand access to titles, tags, and other metadata, and provide contextual information about previously viewed items, current keywords, and alternate keyword possibilities.

Categories and Subject Descriptors

General Terms
Human Factors, Design.

Keywords
Tag, folksonomy, visual interface, search, browse, explore, multimedia.

1. INTRODUCTION

Social software applications such as social bookmarking have helped develop a new method for collaboratively creating and managing tags to annotate and categorize online content. These systems are called folksonomies. Using folksonomies, users can organize their own digital collections, and categorize the content of others’ collections, plus build classification systems. The primary objective of a folksonomy is to help users to effectively search, browse, explore, and organize an information collection. For instance, in the science domain, an effective folksonomy might help users to find scientific articles on a specific topic of interest. Similarly, a multimedia folksonomy might help users to find interesting videos on the Internet. By ‘tagging’ items of interest, users simultaneously bookmark those items for later use and contribute to an unstructured classification system that helps others find the items. In order to achieve these goals, the folksonomy’s interface must be usable, effective, and engaging. Usability is particularly important in this domain because the unstructured nature of tags may make searching difficult, particularly when users are unaware of the index terms used by others.

Many diverse interfaces for folksonomies have recently appeared. Many of these interfaces are designed to grab users’ attention by including catchy and unique visual metaphors. Perhaps this is because designers anticipate that users will be attracted and retained based on these features. Despite the potential advantages that visual interfaces provide and the fact that many users may try them initially, few users elect to use them on an ongoing basis. Perhaps the focus on aesthetics may come at the cost of poor usability and utility. In order to design a visual interface that will elicit continual use, we first need to establish a better understanding of users’ goals and strategies. In the past, researchers typically tested the usability of a folksonomy according to task completion time, mistakes, and subjective evaluations of satisfaction [6]. However, these indices cannot explain the types of search tasks that users do, and the types of strategies that users follow when interacting with different visual interfaces.

In this paper, we present a qualitative, exploratory study to explore how users interact with folksonomy interfaces, particularly focusing on how the visual representations of data affect interactions performed during different tasks. More precisely, we investigate how users interacted with three image search interfaces and three video search interfaces, in order to identify search strategies and the features that best support them. We chose to focus on image and video content because searching pictures and videos on the Internet has become popular with social network users; however, we believe some of our results are also relevant to other types of search. We specifically aimed to answer the following questions:
1. How can search tasks be categorized?
2. For different types of search tasks, what strategies do users employ, and how are these strategies affected by interface design?

2. RELATED WORK
In this section, we begin by summarizing the variety of folksonomy interfaces that are currently available. In our study, we chose a small number of interfaces that maximized the variety of these currently available features. We then discuss techniques that have been proposed to improve web search performance and explain why studying users’ search goals may provide additional benefit. Finally we focus more specifically on user search goals, the target of our study.

2.1 Folksonomy Search Interfaces
Numerous search interfaces have been proposed for folksonomies. Tag clouds have been the most popular tag representation and many social software applications have been using them (e.g., Del.icio.us). Cloudalicious built on tag clouds by displaying how a tag cloud (and the underlying folksonomy) develops over time [12]. In addition, diagonal lines drawn between elements allow users to find a pattern that can be interpreted as a collective opinion of different users. Similarly, Dubinko et al. [3] used ‘river’ and ‘waterfall’ metaphors to visualize tags. YouTube (http://www.youtube.com/) designed an animated representation that uses a Bubble-Up metaphor where users can see their selected video turned into a bubble, surrounded by related video bubbles. Many other visual interfaces have been developed, often including features such as visual metaphors and animation. However, it is unclear which of these features actually support users’ search tasks and promote long-term adoption.

2.2 Improving Search Performance
Many techniques have been developed to improve search performance in both folksonomies and more general web content. Many systems have provided automated keyword suggestions that people can use to tag their content. The assumption is that keyword suggestions will increase the consistency in keyword choice across multiple users, thus making searching easier once a common tag term is known. For instance, Elliot and Özsoyoglu [4] used known relationships between users in a social network to improve keyword suggestions. For example, the word “Grandma” could be ascribed to a certain user’s name if the relationship between the users was known. They found that this approach led to more consistent keyword choices, and that searching for photos of certain people by their relationship to the user was faster than a direct keyword search. Another study compared storyboards and multiple views for exploratory search of videos [1]. They found that storyboards work very well as navigation tools; however, when the user needs information that cannot be captured in snapshots, other visual interfaces such as scatterplots, choropleth maps, and dynamic query preview histograms were also useful for video retrieval.

Nauman and Khan [8] presented the idea of using known user preferences in order to generate more relevant search results (personalized web search). They showed that collaborative tagging improved search performance since search results could be tailored to user preferences as implied by their search and access history. Similarly, Hassan-Montero and Herrero-Solana [7] designed a different tag-cloud layout to improve information retrieval, based on clustering of similar tags. They also proposed the use of an alphabetical order tag cloud with the aim of improving browsing experience. Although these studies may have improved search performance, they did not try to understand and support different types of user goals and strategies. Understanding these goals and strategies may identify unexpected design guidelines and may allow interface customization that could further improve performance.

2.3 Users’ Search Goals
User goals and attention have been studied to some degree. For example, in an eyetracking study, Sutcliffe and Naumone [13] found that fixation densities were related to users’ level of interest in different parts of a website. More similar to our work, Rose and Levinson [11] proposed a framework that categorizes users’ search goals as Navigational, Informational or Resource goals. Informational searches focus on obtaining information about the search topic. For example, the user may wish to answer a question or find advice about a topic. Resource searches represent a goal of obtaining something other than information ‘about’ a topic. This includes obtaining artifacts such as song lyrics, downloading software or other items, as well as obtaining entertainment. Finally, in a Navigational search, the user wishes to be taken to a specific web location such as the home page of a specific university. In our study, all the goals we observed could be considered primarily Resource goals. Our results provide a break-down of the types of tasks users perform when their goal is to find a multimedia resource. In addition, we identify common strategies that users employ and describe how these strategies are influenced by different design features.

3. METHOD
We conducted a qualitative exploratory study to observe how users interact with multimedia search interfaces. We examined interfaces with several different design styles in order to explore factors that affect tasks, interaction strategies, and motivation. Within this context, we assigned participants both closed and open-ended search tasks in order to model the range of multimedia search tasks performed by the average user under normal circumstances.

3.1 Interfaces
Five interactive web-based interfaces for folksonomies were chosen for our study. YouTube and Flickr (http://www.flickr.com/) are two popular sources for searching videos and images respectively. They also allow users to tag, rate and comment on content. YouTube and Flickr provide an Application Programming Interface (API) for searching and accessing their content, which enables other developers to design different interfaces on top of their databases. We chose three such interfaces, with very different design and interaction styles: PicLens (a Firefox web browser plug-in), TagGalaxy (http://taggalaxy.de/) and YouTube Bubble-Up. Example screenshots of these five interfaces are shown in Figures 1-5.
YouTube and Flickr websites allowed users to search, browse, and explore site content in a traditional way familiar to most Internet users. With TagGalaxy, users were allowed to input only one keyword (tag) in the search box to start. In the next level, a “3D galaxy” metaphor was used to represent the most related tags which were shown as planets. Clicking a planet filtered the search results with the additional tag and showed the galaxy view again with new set of related tags. Clicking the sun showed thumbnail images on the surface of the sun and meanwhile enlarged the sun itself. Users could scroll the sun to search through the thumbnails and enlarge an image by clicking its thumbnail. PicLens allowed users to search with multiple tags in the search box. Image or video results were shown by using a ‘gallery’ metaphor with a slide show design and both 2D and 3D styles were optional. Users could scroll the search results horizontally and click a thumbnail to enlarge the image or watch the video. YouTube Bubble-Up used “bubbles” as a metaphor to represent video results after a user performed a search at the YouTube site. Moving the mouse over the a bubble gave the user information about the video and popped up other related video bubbles. A line and different colored dots were used to keep track of videos that had been watched (i.e. a history, as shown in top left of Figure 5).
Table 1: Assessment of differences between the five folksonomy search interfaces.

<table>
<thead>
<tr>
<th></th>
<th>PicLens</th>
<th>Flickr</th>
<th>YouTube</th>
<th>TagGalaxy</th>
<th>Bubble-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metaphor</strong></td>
<td>Gallery</td>
<td>None</td>
<td>None</td>
<td>3D Galaxy</td>
<td>Bubbles</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td>Horizontal scroll</td>
<td>Menu, links, vertical scroll</td>
<td>Menu, links, vertical scroll</td>
<td>3D scroll (i.e. rotate object)</td>
<td>Horizontal &amp; vertical scroll, select adds related items</td>
</tr>
<tr>
<td><strong>Background color</strong></td>
<td>Black</td>
<td>White</td>
<td>White</td>
<td>Black</td>
<td>Black</td>
</tr>
<tr>
<td><strong>Layout presentation</strong></td>
<td>Infinite rows, 2D, 3D</td>
<td>Conventional block layout, 2D</td>
<td>Conventional block layout, 2D</td>
<td>Sphere layout, 3D</td>
<td>Node-link, 2D</td>
</tr>
<tr>
<td><strong>Use of media</strong></td>
<td>Animation, Text shows up when selecting image or video</td>
<td>Text and still images</td>
<td>Text and still images</td>
<td>Animation, Text shows up when selecting image or video</td>
<td>Animation, Text shows up when selecting image or video</td>
</tr>
<tr>
<td><strong>Search type</strong></td>
<td>Multi-tag search</td>
<td>Multi-tag search with keyword suggestion</td>
<td>Multi-tag search with keyword suggestion</td>
<td>Single tag search</td>
<td>Initial video from YouTube search (no refinement), select adds new items</td>
</tr>
<tr>
<td><strong>Snapshots per scene</strong></td>
<td>~36</td>
<td>~10</td>
<td>~10</td>
<td>~50</td>
<td>~30</td>
</tr>
<tr>
<td><strong>Related content</strong></td>
<td>Not available</td>
<td>Separate section</td>
<td>Separate section</td>
<td>Not available</td>
<td>Pop up as bubbles</td>
</tr>
<tr>
<td><strong>Social factor</strong></td>
<td>Not available</td>
<td>Comments</td>
<td>Comments and Ratings</td>
<td>Not available</td>
<td>Ratings</td>
</tr>
</tbody>
</table>

Key differences between the five interface designs are summarized in Table 1. These are broken down by factors that we expected to affect user interaction and engagement. This list of factors extends work by Sutcliffe et al. [13].

### 3.2 Participants and Apparatus

Thirteen paid participants (7 female, 6 male) were recruited from various disciplines at our university. Participant age ranged from 22 to 35. All participants had 3 years or more Internet experience and most used the Internet 20-30 hours per week. They were all familiar with at least one of the image or video search websites examined in our study. Some participants were native English speakers (5 participants) and some were fluent non-native speakers (8 participants).

Our study was run using an AMD Athlon 64 bit dual core PC running at 2.71GHz, with 2GB of RAM, and Windows XP. Both primary and second displays were 22” LCDs at 1650 x 1024 resolution. Participants interacted with the computer using a standard keyboard and mouse.

### 3.3 Tasks

The study had two sections: a picture section involving three tasks per interface, and a video section with one task per interface (see Table 2).

Table 2: Tasks assigned to participants in our study

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pictures</td>
<td>Find five pictures of a famous place shown on an example picture</td>
</tr>
<tr>
<td></td>
<td>Find five pictures that contain similar objects to an example picture</td>
</tr>
<tr>
<td></td>
<td>Find five pictures of different tourist places that you would like to visit. First choose a destination country from a list of options.</td>
</tr>
<tr>
<td>Videos</td>
<td>Find videos that are interesting to you, and watch part of at least 5 videos in a ten-minute interval.</td>
</tr>
</tbody>
</table>

In the picture section, participants were told to try to find the pictures shown during the tasks, but they were not required to
find the exact picture. They were also told to spend a maximum of three minutes on each picture task since we were not interested in accuracy but rather in interaction with the visualizations. In the video section, participants could search for any video they were interested in watching since we wanted them to interact with the visualization at their leisure. Participants used Flickr, PicLens, and TagGalaxy for image search, and YouTube, PicLens, and YouTube Bubble-up for video search. All participants tried all interfaces.

3.4 Procedure
We observed, video recorded, captured the screen, and took notes while participants interacted with the interfaces. Each session was approximately 90 minutes. Participants first completed a background questionnaire. Next they completed a short tutorial on the interfaces, tasks, and scenarios used in the study, followed by a sample task. Once it was clear how to interact with the interface, each task was given, in turn, and participants were asked to spend about three minutes on each picture task and ten minutes on each video task. Task descriptions were shown on the second monitor and search interfaces on the primary monitor. A timer on the second monitor could be used to estimate the time. Upon completing the picture section, participants were interviewed about their experience. After completing the video section, participants were interviewed about their problems, experience, preference and recommendations.

4. RESULTS
We analyzed our data using a qualitative approach inspired by grounded theory [5]. Video interview, video screen capture, and interview notes were analyzed using an open coding approach described [2] to form initial coding categories of strategies used in different types of tasks. Initial coding categories were merged into more general categories in subsequent passes. Here we present our most salient findings.

4.1 Types of Search Tasks
Based on our observations, we identified four different types of search tasks. We categorized search tasks according to whether or not users knew the target of the search and keywords that would identify the target. The following table describes the four search tasks indentified.

These categories of search are differentiated by the type of information available to the user to help them accomplish a search goal. This is in contrast to Rose and Levinson’s [11] categorization of search goals (i.e. user’s reasons for performing the search). In our study, the users’ goals were always Resource in nature. In other words, the user’s goal was to find image(s) or video(s), which are resources.

These categories were derived primarily from the picture search tasks, but we believe that they likely also apply to video search. Video tasks that we observed fell mostly into one category: Search Unknown Objects + Known Keywords. Other categories would likely have been observed if we had assigned more specific search tasks. We found that the information available to a user strongly influenced the user’s strategy and the types of features that were helpful to the search task. These ideas will be explored in the next section.

4.2 Strategies
Here we identify common strategies employed by users. Only successful strategies are reported since the majority of tasks were completed successfully.

<table>
<thead>
<tr>
<th>Known Objects (KO)</th>
<th>Unknown Objects (UO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KK+KO Example</td>
<td>KK+UO Examples</td>
</tr>
<tr>
<td>Find a picture of the Great Wall. Some participants already knew the place and the exact name.</td>
<td>Search for pictures of a frog on a leaf or pictures of pink flowers. In this case, either the frog or the flowers could be any kind.</td>
</tr>
<tr>
<td>UK+KO Example</td>
<td>UK+UO Example</td>
</tr>
<tr>
<td>Find a picture of a pyramid in Egypt. Some participants knew neither the exact place name nor related keywords.</td>
<td>Search for pictures of the Manchu Picchu ruins based on a photo. Some participants were not familiar with the ruin nor the place.</td>
</tr>
</tbody>
</table>

Table 4: Strategies for Known Objects / Known Keywords

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Example</th>
<th>Features that Support Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter keyword(s), then scan a gallery</td>
<td>PicLens presented results using a “gallery” metaphor that allowed users to see many pictures at once. Six participants mentioned that this feature helped them scan pictures and videos quickly. One user said: “I do not read the comments on pictures or videos since I already know what I am looking for, so I just focus on the images.”</td>
<td>Ability to enter multiple keywords</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Many images at once</td>
</tr>
</tbody>
</table>

4.2.1 Known Objects / Known Keywords
When searching for Known Objects / Known Keywords, we observed that users consistently approached the problem by typing in the known keywords and then scanning the resulting image set. This typical strategy and examples are given in Table 4. One feature that was important to support this strategy was the ability to enter multiple keywords. For example, TagGalaxy allowed users to type only one keyword, and then offered alternative keywords, but selecting one of these keywords initiated a new search rather than a refinement. For example, if a user typed “France”, then TagGalaxy gave options such as “Paris”, “Church”, “Tower”, “Sky”, and

---

HCI 2009 – People and Computers XXIII – Celebrating people and technology 414
An Unknown Objects / Known Keywords search task occurred when users did not know their target (e.g., Machu Picchu Ruins), but they had some keywords that helped them find their target (e.g., “ruins”). Table 5 describes three common strategies that were indentified when participants performed a (UO + KK) search. Our key finding was that participants started with keywords they knew and then gradually refined their search; this approach was easiest when the system suggested keywords. Either explicit keyword suggestions or implicit keywords (attached to other images) could be used, but explicit suggestions were generally preferred.

It is also worth noting that meta-data (titles and tags) were used in unexpected ways. Seven participants reported that they did not read the metadata associated with images or video in any detail. However, as noted in the table 5, many of these participants skimmed the metadata to identify keywords to refine their search.

### 4.2.3 Known Objects / Unknown Keywords

A Known Objects / Unknown Keywords search occurred when users had great knowledge about what they were looking for, but did not know how to spell their keywords, or did not know the exact name of the target. For example, four users knew about the “The Great Wall”, but since their first language was not English, they did not know the exact English name. Several other participants knew about the “Machu Picchu” ruins or the “Eiffel Tower”, but did not know the spelling. Table 6 describes the three types of strategies that were indentified when participants performed a (KO + UK) search, and the types of features that supported these strategies.

### Table 6: Strategies for Known Objects / Unknown Keywords

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Example</th>
<th>Features that Support Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter one keyword, then refine</td>
<td>When searching for pictures of Machu Picchu ruins, many participants</td>
<td>Keyword suggestions</td>
</tr>
<tr>
<td>using suggested keywords</td>
<td>typed “ruins”, and TagGalaxy gave them options such as “Temple”, “Rome”, and “Mexico” to narrow their search. Participants followed these suggestions until they felt comfortable with their results. Seven participants reported that this suggestion feature was helpful.</td>
<td>Ability to enter multiple keywords</td>
</tr>
<tr>
<td></td>
<td>Five participants typed a keyword and then went through the results to find pictures similar to their target. When they found similar pictures, they read the title to get ideas that could help them narrow their search.</td>
<td>Title of picture</td>
</tr>
<tr>
<td></td>
<td>Some participants used a strategy similar to the previous one, but learned keywords from image tags. After typing keyword(s) and scanning the results for similar images, they read the image tags and used relevant keywords in their next search attempt. Five participants employed this strategy using Flickr.</td>
<td>Ability to enter multiple keywords</td>
</tr>
<tr>
<td></td>
<td>List of tags for each picture</td>
<td>Many images at once</td>
</tr>
<tr>
<td></td>
<td>Ability to enter multiple keywords</td>
<td>Clear start &amp; end to image set</td>
</tr>
<tr>
<td></td>
<td>Many images at once</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear start &amp; end to image set</td>
<td></td>
</tr>
</tbody>
</table>

### 4.2.2 Unknown Objects / Known Keywords

When searching for pictures of Machu Picchu ruins, many participants typed “ruins”, and TagGalaxy gave them options such as “Temple”, “Rome”, and “Mexico” to narrow their search. Participants followed these suggestions until they felt comfortable with their results. Seven participants reported that this suggestion feature was helpful.

Enter one keyword, scan images, then refine with new keywords from titles of similar images

Five participants typed a keyword and then went through the results to find pictures similar to their target. When they found similar pictures, they read the title to get ideas that could help them narrow their search.

Enter one keyword, scan images, then refine with tags from similar images

Some participants used a strategy similar to the previous one, but learned keywords from image tags. After typing keyword(s) and scanning the results for similar images, they read the image tags and used relevant keywords in their next search attempt. Five participants employed this strategy using Flickr.

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Example</th>
<th>Features that Support Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter one keyword, then refine</td>
<td>When searching for pictures of Machu Picchu ruins, many participants</td>
<td>Keyword suggestions</td>
</tr>
<tr>
<td>using suggested keywords</td>
<td>typed “ruins”, and TagGalaxy gave them options such as “Temple”, “Rome”, and “Mexico” to narrow their search. Participants followed these suggestions until they felt comfortable with their results. Seven participants reported that this suggestion feature was helpful.</td>
<td>Ability to enter multiple keywords</td>
</tr>
<tr>
<td></td>
<td>Five participants typed a keyword and then went through the results to find pictures similar to their target. When they found similar pictures, they read the title to get ideas that could help them narrow their search.</td>
<td>Title of picture</td>
</tr>
<tr>
<td></td>
<td>Some participants used a strategy similar to the previous one, but learned keywords from image tags. After typing keyword(s) and scanning the results for similar images, they read the image tags and used relevant keywords in their next search attempt. Five participants employed this strategy using Flickr.</td>
<td>Ability to enter multiple keywords</td>
</tr>
<tr>
<td></td>
<td>List of tags for each picture</td>
<td>Many images at once</td>
</tr>
<tr>
<td></td>
<td>Ability to enter multiple keywords</td>
<td>Clear start &amp; end to image set</td>
</tr>
<tr>
<td></td>
<td>Many images at once</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear start &amp; end to image set</td>
<td></td>
</tr>
</tbody>
</table>
TagGalaxy did not offer spelling suggestions, so participants had a hard time performing this type of task. During the interview, at least four participants suggested that TagGalaxy should implement a spelling suggestion feature. For a Known Object / Unknown Keyword search task, spelling was the most common problem; therefore, “Spelling Suggestions” was a key feature.

4.2.4 Unknown Objects / Unknown Keywords

Unknown Objects / Unknown Keywords search tasks occurred when users knew little about their target and did not know keywords to describe their target. In this case, participants had to resort to a roundabout strategy as described in Table 7.

This type of task was particularly difficult, and users needed as much help as they could get from the interface. Interfaces with fewer of the features described above (e.g., TagGalaxy) made the task more challenging. We therefore believe that all of the features described in Tables 4-6 would be helpful for a UO + UK search.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Example</th>
<th>Features that Support Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guess keywords, look for pictures containing similar objects, then refine search with new keywords from image titles or tags</td>
<td>Five participants did not know ‘llamas’ nor ‘Machu Picchu’ ruins, which were presented together. They guessed keywords to find pictures containing animals similar to llamas. Then, they copied the keyword ‘llama’. Sometimes those pictures had titles such as “llamas in Machu Picchu”, “Peru—llama”. Participants used these new keywords to refine their search.</td>
<td>All features described above</td>
</tr>
</tbody>
</table>

4.3 Other Observations

Although it was not the focus of our study, we made some interesting observations regarding users’ motivation. We noticed that users were impressed when first introduced to TagGalaxy, PicLens and the Bubble-Up visualization. Users’ first words when introduced with these visualizations were typically: “That is cool”, “That looks nice”, and “that looks fun”. So, 3D graphics and animation seemed to successfully catch the attention of users at first. However, after interacting with such visualizations, opinions often changed.

For example, during the course of the study one of the participants said: “I do not need fancy stuff to search for pictures; all I need is an application that I can use easily”. By ‘fancy’, the participant was referring to features such as 3D graphics and animation. The participant also mentioned that if he was going to use fancy applications, he would use them when he had some time to kill. A different participant said, “When I try to find pictures or videos, what it is more important to me, is that I can see more pictures since I focus more on the pictures or snapshots or videos rather than comments”.

We also observed that motion was not always effective. Participants often spent more time with TagGalaxy, because they rotated the 3D sphere with image snapshots repeatedly in order to search for pictures. Motion was also distracting in the Bubble-up visualization. Motion caused by selecting a new bubble caused disorientation; participants could not find the previous video that they were watching. These observations indicate that positive experiences with multimedia search interfaces, and subsequent long-term adoption, depend more on users’ ability to accomplish search tasks than on aesthetic appeal. This work agrees with Norman’s [9] suggestions that aesthetic design and usability should have a balance. An interface that is less aesthetic is not better than one that has more aesthetic attributes if both lack usability.

5. DISCUSSION: DESIGN FEATURES AND LESSONS LEARNED

Our study categorizes multimedia search tasks, and a series of corresponding strategies employed by users in order to perform these tasks. Specifically, we identified four distinct types of search tasks: Search Known Objects + Known Keywords, Search Known Objects + Unknown Keywords, Search Unknown Objects + Known Keywords, and Search Unknown Objects + Unknown Keywords. An ideal search interface would support these four types of search by including the features identified in tables 4-7. However, not a single visual interface in our study included all of these features.

Even for simple search tasks (KK + KO), simple features such as a clear start and end to the image set, a large ‘gallery’ of images, and the ability to enter more than one keyword helped people to search more efficiently. For more difficult search tasks, where keywords and / or target objects were unknown, additional features were also important. Displaying titles and tags enabled users to identify better keywords to refine their search. Interfaces that offered spelling and keyword suggestions were similarly helpful, since ‘guessing’ keywords was a great frustration for users. Implementing even a few of these features in combination would help users perform multimedia searches more efficiently.

One problem with many of the interfaces we studied is that they did not provide sufficient context information. With TagGalaxy and YouTube Bubble-up, users became spatially disoriented as the view changed (i.e. as the 3D sphere spun around or as new bubbles were added to the scene). Adding visual landmarks might help to reduce these problems. For example, the current object could be more clearly highlighted in the Bubble-up visualization. Search interfaces also provided insufficient context about tags and keywords. For instance, TagGalaxy did not show image titles or all of the tags associated with each image. This lack of information made it difficult for users to learn better keywords based on items similar to their target.

The findings described above may present visual interface designers with somewhat of a dilemma. On the one hand, we found that users wanted to rapidly scan as many images as possible, without taking up screen space for text. An interface that makes users read a lot of text could be considered inefficient, and users might be inclined to discontinue using it. On the other hand, we found that text associated with images (especially tags and titles) enabled users to better refine their search queries. A compromise might be to focus on image
presentation, but provide associated text on-demand (e.g. via small fonts or lightweight actions such as hover queries).

We observed that people cared more about the ability to perform tasks effectively than about aesthetic attributes of the interface. The appeal of such aesthetic attributes is what Aza Raskin [10] has called “interaction seduction.” Although aesthetic appeal can be useful to attract users initially, our study indicated that the importance of visual appeal is superseded by the importance of utility. That is, users are most satisfied by an interface design that allows them to understand features well enough to perform their search tasks efficiently. It is worth noting that this general principle could apply to interfaces other than search interfaces as well. We also observed that visual interfaces with less animation often offered more content or information that helped participants perform their search tasks faster. This does not mean that animation should be avoided altogether, but we suggest evaluating its level of distraction and ensuring that it does not reduce information content.

6. CONCLUSION AND FUTURE WORK

The exploratory study presented in this paper provides a useful starting point for future investigations of users’ experiences interacting with visual interfaces for multimedia search. Our results reveal how users perform different types of searches depending on their knowledge of the search target and their knowledge of keywords to reach that target. We demonstrate how different search strategies are employed depending on the type of search task being performed. We also identify design features that help users perform these particular tasks. Based on our results, we recommend that image and video search interfaces maximize screen space used for visual representations of the media, allow meta-data to be accessed on-demand, and provide contextual information about previously viewed items, current keywords, and alternate keyword possibilities.

Our study focused on search for resources, specifically images and videos. Future work could examine the extent to which similar strategies are used to accomplish navigational and informational search goals, and when searching other types of content such as medical databases and personal files. In addition, long-term studies could examine factors affecting interface adoption in greater detail, and logging or journal studies could examine the interfaces in a more realistic context. We plan to use our findings to prototype better multimedia search interfaces. We also plan to further explore the relationship between search tasks, strategies, design features, and users’ motivation for using visual interfaces. Studies like ours provide the foundation for future visual interface design by offering a heightened understanding of the motivations and needs of users, allowing us to make informed recommendations for design.

7. ACKNOWLEDGMENTS

We thank Nando De Freitas for helping us to initiate this work, and the members of the UVic VisID lab for their helpful suggestions. This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

8. REFERENCES


