What Do Exploratory Searchers Look at in a Faceted Search Interface?

Bill Kules[1], Robert Capra[2], Matthew Banta[1], and Tito Sierra[3]
kules@cua.edu, rcapra3@unc.edu, matt.banta@gmail.com, tito_sierra@ncsu.edu

[1]The Catholic University of America
School of Library and Information Science
Washington, DC

[2]University of North Carolina
School of Information and Library Science
Chapel Hill, NC

[3]North Carolina State University
NCSU Libraries
Raleigh, NC

ABSTRACT
This study examined how searchers interacted with a web-based, faceted library catalog when conducting exploratory searches. It applied eye tracking, stimulated recall interviews, and direct observation to investigate important aspects of gaze behavior in a faceted search interface: what components of the interface searchers looked at, for how long, and in what order. It yielded empirical data that will be useful for both practitioners (e.g., for improving search interface designs), and researchers (e.g., to inform models of search behavior). Results of the study show that participants spent about 50 seconds per task looking at (fixating on) the results, about 25 seconds looking at the facets, and only about 6 seconds looking at the query itself. These findings suggest that facets played an important role in the exploratory search process.

Categories and Subject Descriptors
H5.2 User Interfaces: Evaluation/methodology; H3.3 Information Search and Retrieval

General Terms
Experimentation, Human Factors, Measurement

1. INTRODUCTION
Faceted search and categorized overviews are becoming accepted techniques to support complex information seeking tasks like exploratory search. They use structured metadata to provide an overview of results and incorporate clickable categories into search results. This allows searchers to narrow and browse the results without reformulating their queries. There are a growing number of applications that use these techniques for library catalogs, web search, online shopping, image collections, and other domains [11, 25, 31]. Of particular interest to the library and information science (LIS) field is the application of faceted search interfaces to library online public access catalogs (OPACs) [1]. Libraries invest precious time and money into the development of bibliographic data, and faceted interfaces are a promising way to leverage that investment, making controlled vocabulary visible in a context-appropriate manner at the patron's point of need – their search. By implementing faceted search interfaces, libraries are responding to the demands of a generation of patrons that came of age with web search engines [18].

Faceted techniques have the potential to support more flexible information seeking strategies by allowing searchers to fluidly transition between browsing and search strategies, providing an alternative to query reformulation. Moreover, searchers with partially defined or evolving information needs can use the overview to inform their understanding of a knowledge domain and refine their information needs. Searcher actions, tactics and strategies are affected by the capabilities made available in the search system [2, 13] and therefore by the availability and presentation of such faceted interfaces. For example, use of a faceted search interface can reduce query reformulation and may influence searcher tactics as well as the results searchers choose to examine [17]. Design guidelines for faceted interfaces are beginning to emerge [10, 17], but a better understanding of how they affect searcher actions and tactics would benefit designers.

The exploratory search community has been interested for some time in how people make use of facets in the exploratory search process beyond what can be observed from query refinement and click data. This study addresses such questions by examining gaze behavior to learn what parts of the faceted interfaces searchers attend to, for how long, and in what order. This is the first study we know of that examines how people use facets by analyzing gaze data. This is an important perspective on user behavior that complements other approaches. The results from this exploratory study contribute to our understanding of the role of facets in exploratory search and provide groundwork for future studies that will test specific hypotheses.

JCDL'09, June 15–19, 2009, Austin, Texas, USA.

2. RELATED WORK

The section briefly reviews literature in three areas relevant to our study: faceted search interfaces, eye tracking studies of search interfaces, and exploratory search.

There is a growing body of research on faceted search interfaces. The Flamenco system [11, 31] provided interfaces to specialized collections (art, architecture and tobacco documents), using faceted hierarchies to produce menus of choices for navigational searching. A usability study compared the interface to a keyword-based search interface for an art and architecture database for structured and open-ended, exploratory tasks. With Flamenco, users were more successful at finding relevant images (for the structured tasks) and reported higher subjective measures (for both the structured and exploratory tasks). The exploratory tasks were evaluated using subjective measures, because there was no (single) correct answer and the goal was not necessarily to optimize a quantitative measure such as task duration.

Research on faceted search interfaces has often used two methods to study faceted search interfaces: large-scale log analysis and comparative user studies. Log analysis uses search engine log data to determine what interface elements searchers click on. User studies have employed small-scale log analysis, task completion and satisfaction measures, and concurrent or retrospective verbal reports to discern how searchers used the interface. In previous research that applied these techniques [17], results showed that when using a faceted search interface subjects explored their search results more broadly than without facets while feeling more organized about their searches. Subjects also felt that the faceted interface helped them better assess their results. The verbal reports also suggested seven tactics that participants began to adopt when using a faceted interface. Other faceted search studies have used similar techniques [6, 31]. Faceted search interfaces are not always preferred by users, even when they offer comparable effectiveness [6]. A limitation of these methods is that they do not provide direct evidence of what elements of an interface searchers look at. This study begins to fill that gap by the use of eye tracking analysis.

Eye tracking is an attractive technique for studying web search because gaze-related behaviors are interpreted as indicators of cognitive activity [14]. Fixations (when the eye is stable for a minimum period of time, usually 100-200 msec), gaze duration (the cumulative duration of fixations), and scanpaths (the spatial arrangement of a sequence of fixations) have been used as indicators of attention and cognitive processing in HCI studies. Eye tracking has been used successfully to examine how users interact with search results [8, 9, 19, 23] and to evaluate specific design elements of a prototype web portal application [8]. In [7], the authors examined the effect of varied snippet lengths in search result pages, using gaze data to determine the order in which searchers examined results and the amount of time spent examining individual results. They concluded that longer snippets enhanced performance on information tasks but reduced performance on navigational tasks. Despite the demonstrated value of eye tracking methods for studying non-faceted web search interfaces, we are aware of few studies that use eye tracking to study faceted search interfaces in detail. One study of a faceted search interface for a library catalog has incorporated a limited eye tracking protocol to identify interface elements looked at, but without examining detailed behavior [32].

Exploratory search tasks arise from situations where users "lack the knowledge or contextual awareness to formulate queries or navigate complex information spaces, the search task requires browsing and exploration, or system indexing of available information is inadequate," and inherently have uncertainty, ambiguity and discovery as common aspects [29]. Exploratory search can incorporate other types of search tasks, such as known item, navigational search and question answering, but there is an emphasis on the information seeking context. The high level goal of the information seeking problem affects how users interpret the tasks, their relevance, and the results [17]. Exploratory tasks can be roughly characterized as either learning-oriented or investigative, thus distinguishing them from lookup-oriented tasks [22]. The stage of search can be characterized in various ways, but they generally identify one or more early stages characterized by uncertainty of goals, sources, topics, and terminology [15, 27]. Task complexity [5] is related to exploratory search because complex tasks are less well structured and the searcher often has less knowledge of the topic and poorer conceptualization of the problem [26]. Problem structure, task complexity and prior knowledge have an interconnecting impact when searching. Exploratory search tasks in a library catalog are may be conducted using subject searches, but we differentiate exploratory tasks because a subject search in a catalog can take place at any stage of the search process, whereas exploratory search describes the high level goal of the task.

Designing exploratory search tasks (task construction) is an important, but challenging, requirement for successfully evaluating exploratory search interfaces (http://www.ils.unc.edu/ISSS/). The field is at an early stage in the development of methodologies for both the evaluation of exploratory search systems and the development of tasks for such evaluation. White et al. [30, p. 2879] identify a similar need in their discussion of open research areas. However, there is some prior work to draw from in this area. Borlund [4] provides advice for constructing simulated scenarios and tasks, including creating a situation that is relevant to the subjects and providing details to engage the subject in the search. Kules & Shneiderman [17] used four simulated work tasks for journalists constructed around an exploratory search task to evaluate a faceted web search interface. Yee et al. [31] used a set of four exploratory tasks based on image retrieval.

3. RESEARCH QUESTIONS

The objectives of this research include a) improving our understanding of how faceted interfaces affect searcher actions and tactics and b) developing and initially evaluating a procedure for creating exploratory search tasks for evaluations of search systems. The research questions investigate specific aspects of these objectives:

- How long do searchers look at the major elements (facets, results, query box, breadcrumbs, etc.) of the interface?
- In what order do searchers look at the major elements in a faceted search interface?
- What characteristics and attributes are important to include in exploratory search tasks used for evaluations?

4. METHODS

4.1 Task Design

It is always challenging to design realistic and representative tasks for HCI studies. Designing tasks to study exploratory search can be especially difficult because of the need to induce an *exploratory* style search rather than a directed style search. The high level goal of the exploratory search has an effect on users' interpretation of the task and the results produced [17], so task design is especially important. A review of prior work on exploratory search and task design [3, 5, 15, 17, 22, 26, 29] suggests the following desirable characteristics of exploratory search tasks: uncertainty, ambiguity, discovery, be an unfamiliar domain for the searcher, provide a low-level of specificity about how to find the information, and be a situation that provides enough imaginative context for the participant to relate and apply the situation. We used a two-step approach in developing tasks for this study, first extracting candidate topics from the NCSU OPAC logs and then inserting and refining the topics into "task templates" that were designed to meet the desirable characteristics described above. An overview of our task design procedure is given below and more details are given in [16].

Topic Extraction from Log Data

First, we grounded the task topics in real-world usage logs of the North Carolina State University (NCSU) Libraries OPAC system that serves on average 7,824 search transactions and 1,087 user sessions per day [20]. We extracted three days of anonymized log data from the NCSU OPAC, including the keyword queries and a list of any facets selected by the searcher. We then looked through the logs for sessions that had characteristics of exploratory search and that included the use of facets. To operationalize this, we looked for sessions in which the user did not find the desired results on their first interaction and then either reformulated the search or interacted with the results. We disregarded instances where the user needed to use a "show more" option to see additional facet values because we wanted to focus on facets visible from the initial results page. For example, from one log file, we observed queries for the search term "British History" with the facets "History" (subject) and "Twentieth Century" (time period). From these log entries, we developed a candidate topic "British History".

Task Templates and Task Refinement

To situate the topic in a broader exploratory search task, we developed a "template scenario" that involved finding multiple items – to which the specific candidate topics could be plugged into. The idea of the template scenario was to situate the participant in a familiar situation in which multiple items would need to be found. Since we recruited participants from a university population, we used a scenario that involved writing a paper for a class. Early formative evaluation revealed that searchers preferred some structure in the task so we included guidance about the number of topics and books to find:

> Imagine that you are taking a class called ____. For this class, you need to write a paper on the topic ____. Use the catalog to find two possible topics for your paper. Find three books for each topic.

Once candidate task scenarios were created, we tested and refined them to: 1) clarify the wording of the scenario, 2) insure that the scenario was not too easy to qualify for use in an exploratory search (i.e. the first 10 results did not answer the task), and 3) make sure that the scenario benefited from using facets (since facet use was a focus of our study). For example, refinement of our "British History" topic resulted in the final topic "the relationship between Great Britain and its Colonies in the Twentieth Century". The scenarios generated by the process were reviewed by library science experts, pilot tested and further refined with three participants.

Using the process described above, we developed four exploratory task scenarios (see A–D below). We also used two known-item task scenarios (E and F) based on a previous NCSU study:

> A. Imagine you are taking a class called "Feminism in the United States". For this class you need to write a research paper on some aspect of the U.S. feminist movement, but have yet to decide on a topic. Use the catalog to find two possible topics for your paper. Then use the catalog to find three books for each topic so that you might make a decision as to which topic to write about.

> B. Your professor wants you to write a paper comparing the textile industry in three countries in three different continents. Use the catalog to find three countries which have a textile industry about which books have been written. Find three books for each country.

> C. Imagine you are taking a class titled "Great Britain and its

Colonies in the Twentieth Century". For this class you need to write a research paper on some aspect of the relationship between Great Britain and its Colonies in the Twentieth Century but you have yet to decide on one. Use the catalog to find two possible topics for your paper. Then use the catalog to find three books for each topic so that you might make a decision as to which topic to write about.

D. You are taking a class called "History of the Olympic Games" for which you need to write a research paper. You have yet to decide on a specific topic for this paper. Use the library catalog to explore possible topics and find two. Then find at least three books for each so that you might make a decision as to which topic to write about.

E. Your professor has suggested that your group begin your project on Conservation and Biological Diversity by looking up background information in a book titled Firefly encyclopedia of trees.

F. You are working your way through the Harry Potter books and are ready to read the next one on your list, titled "Harry Potter and the Goblet of Fire".

4.2 Interface

One of the main goals of this study was to examine how people used facets in an OPAC as part of the exploratory search process. We used a variant of the NCSU OPAC interface that included three facets, support for keyword search, and a main results area (see Figure 1). The three facets (Subject, Region, and Time Period) were included because they are the three most relevant facets for supporting exploratory search available in the NCSU system. The keyword search interface allowed users to search over all the catalog record fields (anywhere), or to narrow the search to a specific field such as title or author using a pull-down menu.

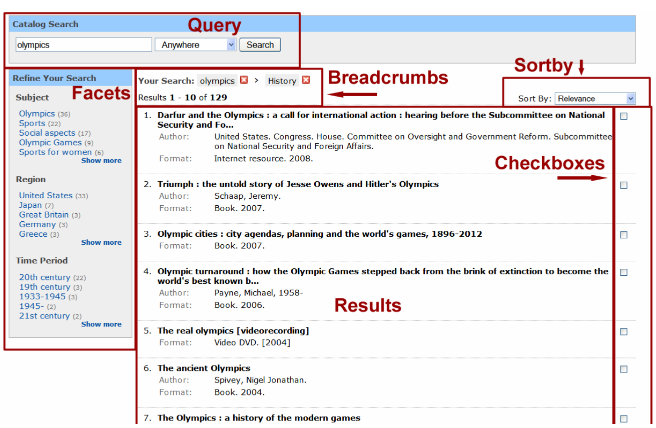

Figure 1. Interface used for the study, showing six (6) areas of interest (AOIs). Navigation and Suggestion AOIs are not shown.

The boxes denote the eye-tracking areas of interest (AOIs) and will be described in more detail in the Data Analysis section of this paper. A check box was displayed next to each item in the results list. Participants were instructed to use this check box to indicate the items they were selecting for each task.

4.3 Materials & Equipment

The interface used for this study is based on the NCSU Catalog Research Testbed [24], a tool for developing and evaluating faceted library catalog interfaces. The data source for the interface is the NCSU Libraries' production catalog of over 1.8 million item records. The interface design was customized to support the specific requirements of the research study. Some of these customizations included a fixed width interface layout to optimize eye-track data collection, the inclusion of check boxes to indicate user item selection, selected facet display, and the removal of NCSU-specific content (e.g. global navigation, library locations) from the interface.

A Tobii 2150 remote eye tracker (http://www.tobii.com) was used. It embeds the remote infrared cameras in a 21" LCD monitor, which samples the user's eyes at a rate of 50 Hz. Resolution was set to 1024x768.

4.4 Procedures and Experimental Design

The study was conducted in the Human-Computer Interaction Lab at the University of Maryland at College Park (UMD). Twenty-one participants were recruited from the UMD community. Of these, data was successfully collected from 18. Half the 18 were male, half female. Six were undergraduates and 11 were graduate students. 16 of the 18 were under 30 years old and 15 indicated that they used a search engine one or more times a day. On a scale from 1=novice to 5=expert, 14 of the 18 self-rated their searching skills as 4 or 5. Thirteen participants noted that they had used a faceted search system (typically a library catalog) before. The study sessions lasted about 1.5 hours and participants received a $20 honorarium for their participation.

The experiment used a 1x2 within-subject design in which all participants used the same interface to complete two types of search tasks, exploratory (n=4) and known-item search (n=2). The exploratory tasks were presented first, followed by the known-item tasks. Within task type, tasks were counterbalanced. The study was conducted with the participants one at a time.

Each participant was greeted by the experimenter, seated at the computer, and asked to read and sign an informed consent form. Then the participant was shown a short (90 second) video that demonstrated the interface. Next, the eye-tracker was calibrated. Participants then completed six searches motivated by the six task scenarios. In between each task, participants responded to a questionnaire with the following five questions about their experience with the task. All responses were given as ratings on a 5-point Likert-type scales (anchors shown in parenthesis):

1. How familiar were you with this subject when you began this task?
   (1 = not familiar at all, 5 = very familiar)
2. How difficult was it to accomplish this task?
   (1 = very difficult, 5 = very easy)
3. I am confident that I fulfilled the task asked of me.
   (1 = strongly disagree, 5 = strongly agree)
4. To what extent did completing this task involve finding a single item versus finding multiple items?
   (1 = single item, 5 = multiple items)
5. To what extent did you change what you were looking for based on the results you found?
   (1 = not at all, 5 = a lot)

These questions were asked mainly to gather descriptive data about the participants' experiences with the tasks.

After subjects completed all six search tasks, we conducted a semi-structured, stimulated recall interview. Participants were shown screen capture videos of two of their searches with their eye-traces overlaid so they could see where they had been looking. While they viewed these videos, the participants were asked to provide a retrospective verbal report [12] of their thoughts as they were searching. While watching the video, the experimenter would ask general questions such as, "What were you thinking about here?" to help elicit the retrospective verbal reports by the participants. Interviews were audio recorded.

At the end of each session, participants performed a card sort to group the six task scenarios according to which tasks they thought were most similar and also filled out a final questionnaire about their background with search systems.

4.5 Data Analysis

The Tobii Clearview analysis software (v2.7.1) was used to segment each page into eight areas of interest (AOIs) based on interface elements: Query box, Facets, Results, Breadcrumbs, Checkboxes, Navigation links, Sortby (a sort selector), and a Suggestion (displayed when possible spelling errors were detected). These AOIs are shown by bounding boxes in Figure 1. Raw gaze data was analyzed to extract fixations with a minimum threshold of 100 ms duration within a radius of 30 pixels inside the defined AOIs. This was used to compute fixation times and to extract scan paths (the order in which each AOI was viewed).

5. RESULTS

In this section, we present the results of analyzing the data collected in the study. First, we discuss the participants' perceptions of the task scenarios used in the study. Then we examine the gaze duration and gaze transitions for individual areas of interest (AOIs) in the user interface.

5.1 Task Scenarios

Perceptions of scenarios

Table 1 shows the averages and standard deviations (in parenthesis) of the participants' perceptions of the exploratory and known item tasks based on the five questions asked after each task. Subjects were slightly more familiar with the known item tasks and found them somewhat easier. They were also slightly more confident that they had accomplished the indicated task. Participants clearly differentiated between the numbers of items that each task required (single vs. multiple). They also changed what they were looking for more for the exploratory tasks.

	Exploratory n=72 avg (stdev)	**Known-item n=36 avg (stdev)**
1. Familiarity	2.6 (1.39)	3.0 (1.80)
2. Difficulty **	4.0 (0.91)	4.9 (0.23)
3. Confidence **	4.2 (0.94)	4.8 (0.80)
4. Single/Multiple **	4.2 (0.92)	1.4 (1.15)
5. Changed goal **	3.3 (1.33)	1.1 (0.4)

Table 1. Overall Perception Ratings (significant difference found between exploratory and known-item at $p<0.001$ using two-tailed T-test with $\alpha = 0.05$)**

Card sorting the scenarios

We asked the subjects to group the scenarios "and put the ones that are the most alike together into groups." Of the 17 subjects who completed this step, all 17 put the two known item tasks (E & F) in their own group. Nine of the subjects grouped tasks A, C, and D together, placing B separately. Three put A, B, C, D all together. The remainder had various grouping of A, B, C, D. When asked about task B, the explanations focused on the geographic nature of the task and the fact that it asks for books instead of topics, as the other three do. We anticipated the strong distinction between exploratory and known-item, but the sub-distinction of tasks within the exploratory set was unexpected and suggests that subjects considered the geographic/topical and books/ideas differences to be important aspects of the nature of the tasks. We interpret these results to indicate that our task design did succeed in creating two main groups of tasks: exploratory (A, B, C, D), and known-item (E and F).

5.2 Gaze Duration

Gaze duration can be defined as the "cumulative duration and average spatial location of a series of consecutive fixations within an area of interest… a fixation occurring outside the area of interest marks the end of the gaze." [14 p.7] and is a commonly used metric in eye-tracking studies. Longer fixations and gaze durations can indicate areas of importance or cognitive difficulty [14 p.9].

In this study, we were not investigating the complexity of interface elements and kept our interface design simple in order to focus on the participant's use of the AOIs (Results, Facets, Query, etc.). We investigated gaze duration for the tasks in this study with the assumption that longer gaze durations indicate greater engagement of the participant with the content of the AOI. For example, if a participant spent time looking at the Results AOI, we assume that they did so as part of their effort to understand the results and their relevance to the current task goal. This is consistent with the interpretation in [7]. The AOIs with the largest durations were the facets, the query box, and the results list, and our analysis of duration focuses on these three.

Task and AOI

Our first analysis of gaze duration looked across all six tasks, the four exploratory (A, B, C, D) and the two known-item tasks (E, F). A two-way ANOVA was conducted with independent variables Task and AOI and dependent variable gaze duration. The results of the ANOVA showed significant effects of Task, AOI, and Task * AOI at alpha = 0.05. Post-hoc analysis[1] showed that tasks E and F had significantly lower total gaze duration than the other tasks. We anticipated that tasks E and F would elicit more directed search behavior that would require less time and average gaze duration to complete. Tasks E and F were included mainly for the purpose of helping to evaluate whether the exploratory search tasks did elicit, exploratory types of behavior in comparison to the known-item tasks.

Our main interest in this study is exploratory search, so we repeated the ANOVA described above for only the four exploratory search tasks (A, B, C, D). This ANOVA (alpha = 0.05) did not detect any significant difference based on the main effect of Task. A significant main effect of AOI was found ($p<0.01$). No interaction effect of Task by AOI was detected. Post-hoc analysis on the main effect of AOI showed significant differences between all three AOIs considered. The Results AOI had significantly longer gaze durations than the Facet AOI and the Facet AOI had significantly longer gaze durations than the Query AOI. Figure 2 shows the average gaze durations for each AOI for each Task.

Stage of the Search Process

Searchers conducting exploratory searches move through different stages as they work through an iterative process to make queries, evaluate results, modify sub-goals, and refine searches. We wanted to investigate how gaze duration on the main AOIs (Results, Facets, Query) differed based on the stage of the search process. This is difficult to study because it is difficult to know exactly what search activity a user is engaging in at a given point in their overall search.

[1] All post-hoc analyses were conducted using the Tukey-Kramer method.

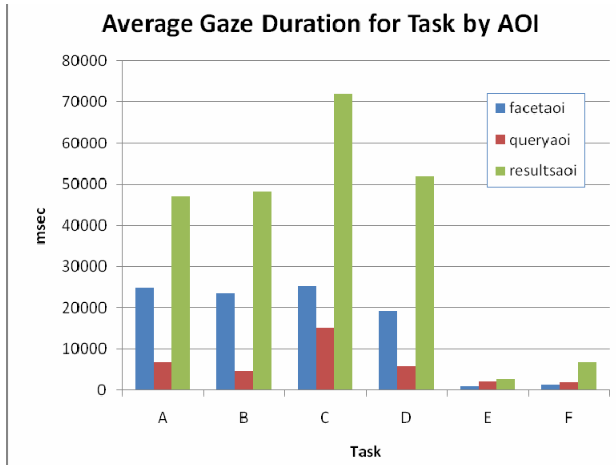

Figure 2. Gaze Duration for Task by selected AOIs.

As a starting point for investigation we looked at the first three pages that the user viewed as they started each task. All participants started at a similar stage when the first page of results was displayed. That is, they had all entered a keyword query and received a set of results and accompanying relevant facets. After this first page of results, participants took a variety of actions, e.g., viewing additional results from the query, narrowing results via the facets, or reformulating their query. They may have diverged as to what stage of the search process they were at, but we thought that patterns might emerge from analysis of the first three pages viewed.

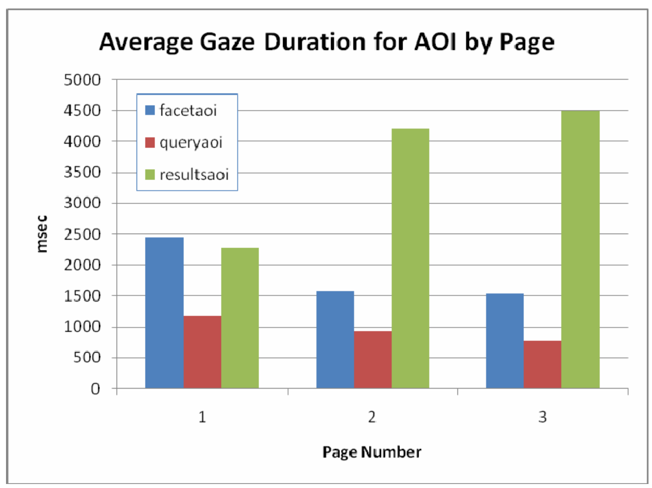

Figure 3. Gaze Duration for selected AOIs by Page Number.

We conducted an ANOVA with independent variables Task, AOI, and Page Number with dependent variable Gaze Duration. Only the first three pages viewed were included in this analysis. Two significant effects were found ($alpha<0.05$): a main effect of AOI and an interaction effect of AOI * Page number. No other significant effects were detected. Analysis of the main effect of AOI is the same as the analysis presented in the previous section. Post-hoc analysis of the AOI * Page number interaction effect showed that in the first page viewed in the search, there was no significant difference between the gaze duration of the Facet AOI, the Query AOI, and the Results AOI. Figure 3 shows the average gaze durations for each AOI for each of the first three pages viewed. A significant difference between the Facet AOI and the Results AOI was found for the second and third pages viewed ($p<0.01$). The Results AOI had a significantly longer gaze duration than the FacetAOI for the second and third pages. A similar significant difference was found between the Query AOI and the Results AOI for the second and third pages.

5.4 Gaze Transition

In addition to examining the relative amount of time participants spent looking at specific AOIs, we also examined the pattern of eye movement across the page. Doing so provides us with a metric for understanding the relationship between different AOIs based on observed gaze behavior.

The eye-tracker software captures a full scan path for each participant page view. For example, if the participant looks at the search query box, then the facet area, and finally the search results, then the sequence path would be: Query → Facets → Results. As outlined in Jacob and Karn [14], we combined sequential fixations within the same AOI into one. After the eye-track sequence data was collected, we processed the sequence path data across all participants and task scenarios to generate a complete list of unique gaze transition pairs for the study. For example, the gaze transition pairs for the earlier example would be: (Query → Facets), (Facets → Results).

Figure 4 illustrates the gaze transitions captured in our study as a directed graph. Each oval represents an AOI. Each arrow represents a gaze transition and suggests a shift in attention. Arrow labels represent the frequency of the specific gaze transition among the entire set of collected gaze transition pairs.

To aid readability, Figure 4 represents only top 10 most frequently occurring gaze transition pairs. Thus, not all arcs between AOI pairs are shown. For example, the arc from Results to Query is shown (4.6%), but the arc from Query to Results is not shown because the transition in that direction did not occur frequently enough be in the top 10 transitions. As shown in Figure 4, the top 10 transition pairs account for 77.2% of all transitions. Three AOIs (Sortby, Suggestion and Navigation) did not generate enough gaze transition activity to appear on this graph at all.

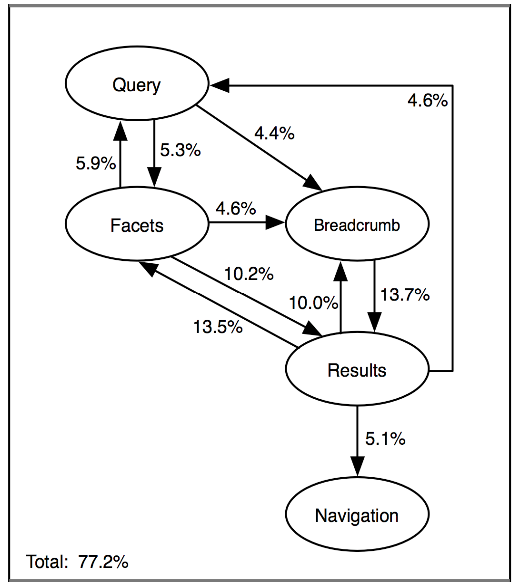

Figure 4. Top 10 gaze transitions between AOIs across all participants, task scenarios, and page views.

Figure 4 indicates that 47.4% of all gaze transitions occurred between just two pairs: the Facet-Results pair and the Results-Breadcrumb pair. The breadcrumbs displayed both the current query and the currently applied facets. The Query-Results pair only accounted for 4.6% of the overall transitions. Overall, facets were involved in 38.5% of all transitions (either to or from the facets).

5.5 Stimulated Recall Interviews

We conducted a preliminary analysis of the post-search stimulated recall interviews to provide insight into how searchers perceived the tasks and the tactics they used to accomplish them. Observations from this analysis are summarized below.

Participants commented on how their attention shifted between the various interface elements. For example, one participant noted that she shifted attention from the results to the facets because she did not understand the subject very well. Another reported finding her topic from the results and then looking through the facets to see if there was anything that corresponded to it, so that she could further narrow her results. She described using the facets as "shopping around", selecting facets and then looking to see what was available for a particular subject. Participants also reported using facets initially to help decide on and narrow a topic (e.g., "I needed a subject and I didn't want to look through 2000 books.").

Many participants mentioned that they found the scenarios to be broad in nature, but that the topics were realistic for undergraduate course papers (e.g. feminism, history of the world, manufacturing). Several participants noted that the broadness of the scenarios required them to find specific sub-topics on which to focus and felt that in real-life situations they might have had a more specific focus from the start. This highlights the delicate balance when constructing exploratory search tasks. The tasks need to be broadly scoped so that they elicit exploratory search behaviors, but also, the participants need to have enough information to ground the tasks.

Several participants mentioned that in "real-life" settings, they would use other sources to help refine topics. For example, one participant noted, that she would use Google or Wikipedia to get topic ideas. Several participants talked about how the facets in the study interface helped fill this role (e.g. "the interface gave me lots of ideas I wouldn't have thought of"). Participants also noted the differences between searching a library catalog and searching the web, (e.g. "this produces pretty broad results… just to give you a sense of what the subject may be about").

6. DISCUSSION

In this study, we set out to investigate three main research questions:

- How long do searchers look at the major elements of the interface?
- In what order do searchers look at the major elements in a faceted search interface?
- What characteristics and attributes are important to include in exploratory search tasks used for evaluations?

We found that participants spent the most time looking at the result items returned in response to their queries. For the tasks in our study, participants spent about 50 seconds per task looking at the results. Interestingly, participants spent about 25 seconds looking at the facets, and only about 6 seconds looking at the query itself. This result is interesting because it suggests that the facets played an important role in the exploratory search process. Participant comments suggest that they used the facets to help organize their view of the topic domain and to select sub-topics for further investigation. At some stages of the search process, participants' interaction with the facets appeared to be as important (or more important) than the results themselves. For example, when trying to decide on the sub-topics for the tasks, participants were often observed spending time reviewing the facets. It is possible that the visual layout influenced searchers – people click on what they see and what is most visible. However, participants did comment that they got ideas from the facets for how to proceed with their search which suggests that the content was an important influence.

One of the most interesting results was the significant effect of AOI by Page number. The results showed that in the first page returned, participants looked at the facets, results, and queries about equally. However, in the second and third pages, the results played a much greater role. This suggests differences in the importance of interface elements such as the facets and results based on the stage that the user is at in the search process. This is consistent with models of the search session (e.g. Marchionini [21]) in which searchers at later points are focused on extracting desired information from search results. In this analysis we only examined the first three pages looked at in the search. In future work we plan to examine the entire space of pages used to see if there are additional differences based on stage of the search process.

The analysis of the gaze transition pairs revealed interesting patterns. First, the facet-result pair and the results-breadcrumb pair accounted for 47.4% of all the transitions. Since the breadcrumb displays both the current query and the currently applied facets, this indicates that looking back and forth between the results and the facets (facets either in the facet area on the left, or in the breadcrumb) is a very important activity in these tasks – one that accounts for almost half of all gaze transitions. Contrasting the transitions to and from the query area with those to and from the facet area, it is clear to see that there were more transitions for the facets. This suggests that facets were heavily used in these tasks (more so than references back to the query itself). In traditional OPAC interfaces, query reformulation is one of the primary methods of query refinement. The faceted interface provided another option that appears to have been very useful. As one participant commented, "The subject thing worked. I don't normally do subject searches."

The use of the facets to help gain an overview of the topic space and to help decide on sub-topics to focus on is a finding that is supported by prior studies of faceted systems. In a study of the Relation Browser interface to find information on the BLS web site, Capra et al. found that "participants listed the use of facets to narrow down a search, and support for gaining an overview of the data as the best features." [6, p.449]. Wilson, et al. in a study of their mSpace faceted browser observed that "facets were used both passively to understand the structure of the collection and actively to produce more expressive queries" [28, p.55]. Our results are consistent with these and suggest that facets can play an important role not only in understanding an information space, but also in helping users select promising sub-topics for further exploration.

Overall, the exploratory tasks achieved our objectives. Based on the participants' perceptions of the tasks, we believe that our procedure for task generation led to well-grounded, realistic tasks that did elicit exploratory search behavior for the exploratory tasks. The exploratory tasks met the desired characteristics we outlined as goals: relatively low initial topic familiarity, require multiple items to be considered, and some ambiguity as to the final answers (as indicated by the confidence and changed goal measures). The difference in scenario B suggests that searchers differentiate between the indicated object (books vs paper topics) and by the nature of the facets (topical vs geographic).

Limitations

One limitation of this study is the use of aggregate fixation duration as a measure of gaze behavior. This provides a useful indicator of attention, but it does conflate gaze frequency with gaze duration. Distinguishing between these two elements would provide separate indications of importance and difficulty of information extraction [14]. The use of researcher-provided tasks based on a single template, the use of only three facets, and the use of a "simplified" OPAC interface limits the generalizability of these results. These constraints were considered necessary to enhance experimental control as we gathered gaze data. Providing training is not typical in the self-service web environment and the results should not be construed as representative of novice users of a faceted search interface.

The procedure for generating and testing the validity of the exploratory search tasks is limited in several important ways. First, the tasks do not explicitly satisfy Borlund's [3] recommendations for making the test topics interesting and relevant to the subjects, because we did not recruit subjects with a particular interest in, for example, history. However, many students are required to write papers on topics that are not of great interest to them. Second, the tasks were explicitly selected to closely match facets provided by the system. This was intentional, because we wished to study gaze behavior when the facets were well-matched to the task. Gaze behavior may differ when the facets are not well-matched to the task. Finally, the operational definition used for exploratory search only accounts for a few of the many possible attributes. The evaluation questions measured several attributes, but other questions could provide additional insight into the validity of the tasks.

7. FUTURE WORK

Future work should investigate additional interface factors and their impact on gaze behavior. These include: the number of facets, whether they extend "below the fold", how well-matched the facets are to the topic/task, searcher domain knowledge and search experience, stage of search, and the high-level work task situation. We currently investigating the provision of short training videos to develop guidelines for library implementers. We will incorporate additional measures such as fixation counts to gain a more fine-grained picture of gaze behavior. We also plan to investigate possible correlations between gaze and click behavior. We are continuing to analyze the post-search interview data to learn how searchers explain their thoughts and actions as they used the interface.

The procedure for generating exploratory search tasks can be generalized beyond the faceted interfaces. Additional investigation and experience will help explicate what additional dimensions of exploratory search tasks contribute to realistic and representative tasks. Additional measures of task perception will help strengthen the ability to assess task validity.

8. CONCLUSION

Facets are becoming more common in search interfaces. Studies have examined the usefulness of facets in exploratory search; however, very few studies have looked specifically at facet use in an OPAC. In addition, little is understood about how users actually use facets outside of clicking on them. For example, do users look at facets as part of an exploratory search for the purpose of formulating and refineing search topics? In this study, we examined gaze behavior to better understand where participants spend their time looking while doing exploratory searches with an OPAC. Our results show that facets played a major role in this process, accounting for about one-half the amount of time spent looking at actual results, underscoring the importance of the facets.

This paper suggests a principled way of task building that incorporates consideration of the dimensions of the task, then building and refining the task description while taking into account both the broader dimensions of exploratory search and the pragmatics of the particular search system and collection technique. We believe that this task development strategy is a step toward making tasks more comparable across studies.

9. ACKNOWLEDGEMENTS

This work was supported in part by grants from the NSF/Library of Congress (ISS 0455970) and NSF grant ISS 0812363 and a grant from the Catholic University Grant-in-Aid Committee. We thank Doug Oard for the use of an eye tracker, and Joseph Ryan and Jason Casden for their help in configuring the interface for this study.

10. REFERENCES

1. Antelman, K., Lynema, E., & Pace, A. (2006). Toward a 21st Century Library Catalog. Information Technology and Libraries, 25(3), 128-139.
2. Bates, M. (1990). Where should the person stop and the information search interface start. Information Processing and Management, 26(5), 575-591.
3. Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. Journal of Documentation, 56(1), 71-90.
4. Borlund, P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. Information Research, 8(3), paper no. 152. Retrieved September 19, 2008, from http://informationr.net/ir/8-3/paper152.html.

5. Byström, K, & Järvelin, K. (1995). Task complexity affects information seeking and use. Information Processing and Management, 31(2), 191-213.

6. Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., & Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. In Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries- (pp. 442-451). New York: ACM Press.

7. Cutrell, E. and Guan, Z. 2007. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 407-416.

8. Goldberg, J., Stimson, M., Lewenstein, M., Scott, N., & Wichansky, A. (2002). Eye tracking in Web search tasks: Design implications. In Proceedings of the 2002 Symposium on Eye Tracking Research and Applications (pp. 51-58). New York: ACM Press.

9. Granka, L., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In Proceedings of the 27th Annual International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 478-479). New York: ACM Press.

10. Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. Paper presented at the ACM SIGIR 2006 Workshop on Faceted Search. Retrieved September 27, 2007, from http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf.

11. Hearst, M., Elliot, A., English, J., Sinha, R., Swearingen, K., & Yee, P. (2002). Finding the flow in web site search. Communications of the ACM, 45(9), 42-49.

12. Henderson, L., & Tallman, J. I. (2006). Stimulated recall methodology. In Stimulated recall and mental models: Tools for teaching and learning computer information literacy (pp. 55-91). Lanham, MD: Scarecrow Press.

13. Käki, M. Findex: Search result categories help users when document rankings fail. In *Proceedings of ACM SIGCHI* (Portland, OR, Apr. 2005).

14. Jacob, R. & and Karn, K. (2003) Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, ed. by J. Hyona, R. Radach, and H. Deubel, pp. 573-605, Amsterdam, Elsevier Science.

15. Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361-371.

16. Kules, B., & Capra, R. (2008) Creating exploratory tasks for a faceted search interface. Paper presented at the Second Workshop on Human-Computer Interaction (HCIR 2008).

17. Kules, B., & Shneiderman, B. (2008). Users can change their web search tactics: Design guidelines for categorized overviews. Information Processing & Management, 44(2), 463-484.

18. Library of Congress (2008). On the Record: Report of the Working Group on the Future of Bibliographic Control. Retrieved September 13, 2008, from http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

19. Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2005). The influence of task and gender on search and evaluation behavior using Google. Information Processing and Management, 42(4), 1123-1131.

20. Lown, C. (2008). A Transaction Log Analysis of NCSU's Faceted Navigation OPAC. Master's paper, University of North Carolina at Chapel Hill, 2008.

21. Marchionini, G. 1997 *Information Seeking in Electronic Environments*. Cambridge University Press.

22. Marchionini, G. (2006) Exploratory search: from finding to understanding. Communications of the ACM 49(4): 41-46.

23. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position and relevance. Journal of Computer-Mediated Communication, 12(3), article 3.

24. Sierra, T., Ryan, J., & Casden, J. (2008). The NCSU catalog research testbed: a tool for evaluating faceted library catalog interfaces. In Proc. of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (p. 467). New York:ACM Press.

25. Tunkelang, D. (2006). Dynamic Category Sets: an approach for faceted search. Paper presented at the ACM SIGIR 2006 Workshop on Faceted Search. Retrieved September 27, 2007, http://www.cs.cmu.edu/~quixote/DynamicCategorySets.pdf.

26. Vakkari, P. (1999). Task complexity, problem structure and information actions - Integrating studies on information seeking and retrieval. Information Processing and Management. 35(6): 819-837.

27. Vakkari, P. (2001). A theory of the task-based information retrieval process: A summary and generalisation of a longitudinal study. Journal of Documentation, 57(1), 44-60.

28. Wilson, M. L. and schraefel, m. 2008. A longitudinal study of exploratory and keyword search. In Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (Pittsburgh PA, PA, USA, June 16 - 20, 2008). JCDL '08. ACM, New York, NY, 52-56.

29. White, R. W., Kules, B., Drucker, S. M., & schraefel, m.c. (2006). Supporting exploratory search. *Communications of the ACM*, 49(4), 36–39.

30. White, R. W., Drucker, S. M., Marchionini, G., Hearst, M., and schraefel, m. c. (2007). Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. In CHI '07 Extended Abstracts on Human Factors in Computing Systems (San Jose, CA, USA, April 28 - May 03, 2007). New York: ACM Press.

31. Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. In Proceedings of the SIGCHI Conference on Human factors in Computing Systems, Ft. Lauderdale, FL (pp. 401-408). New York: ACM Press.

32. Zhang, Y., & Salaba, A. (2007). User research and testing of FRBR prototype systems. Paper presented at the Annual Meeting of the American Society for Information Science and Technology (ASIST).