Search Behaviors in Different Task Types

Jingjing Liu, Michael J. Cole, Chang Liu, Ralf Bierig, Jacek Gwizdka, Nicholas J. Belkin,
Jun Zhang, Xiangmin Zhang
School of Communication and Information, Rutgers University
4 Huntington Street, New Brunswick, NJ 08901, USA
{belkin, m.cole, bierig, jacekg}@rutgers.edu, {jingjing, changl, zhangj}@eden.rutgers.edu, xiangminz@gmail.com

ABSTRACT
Personalization of information retrieval tailors search towards individual users to meet their particular information needs by taking into account information about users and their contexts, often through implicit sources of evidence such as user behaviors. Task types have been shown to influence search behaviors including usefulness judgments. This paper reports on an investigation of user behaviors associated with different task types. Twenty-two undergraduate journalism students participated in a controlled lab experiment, each searching on four tasks which varied on four dimensions: complexity, task product, task goal and task level. Results indicate regular differences associated with different task characteristics in several search behaviors, including task completion time, decision time (the time taken to decide whether a document is useful or not), and eye fixations, etc. We suggest these behaviors can be used as implicit indicators of the user's task type.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *relevance feedback, search process*

General Terms
Design, Experimentation, Human Factors, Measurement, Performance.

Keywords
Personalization, Information retrieval, Task type, User behavior, Eye tracking.

1. INTRODUCTION
A relatively recent, but quite significant approach to improving Web search is *personalization*. Personalization means the tailoring of various aspects of the search experience to specifics of the particular user, including that user's goals, search intentions, individual characteristics, and a variety of other contextual factors. Although "search experience" can encompass a wide variety of aspects of the search, such as adapting the interface to a person's cognitive style, most research and practice has focused on tailoring search results (or advertisements) to the user's specific situation. This has principally taken the approach of attempting to identify the topic (sometimes called "intent") of the person's information problem, and then ranking the results of the search according to that user-specific intent. This is typically accomplished by inferring intent from user behavior, for instance click-through or dwell time, and then applying some form of relevance feedback or result classification.

JCDL'10, June 21-25, 2010, Gold Coast, Queensland, Australia.

One aspect of the information seeker's context that has been shown to affect information seeking behavior is the nature of the task, sometimes called the goal that led the person to engage in information seeking. Extensive studies have addressed the effects of various characteristics of task, including complexity, difficulty, and stage, on search behaviors, including usefulness or relevance judgment (e.g. [5]). Furthermore, White & Kelly [31] have shown that knowledge of task type can improve performance of implicit relevance feedback predicted using dwell time. Their result and similar work (e.g. [12]) motivated our current research.

We are concerned in our research with being able to predict task type based on searcher behavior in the course of an information seeking episode. The rationale for this focus is this: if task type can be predicted from implicit evidence, then that knowledge can be used to interpret implicit indicators of usefulness, relevance, etc., in order to make personalization of the search experience more accurate and useful.

To address this issue, we conducted a detailed study of a group of similar participants, doing realistic and well-defined work and search tasks on the Web, without constraint on the information sources or search engines used. This work reports on several direct measures of the observed search behaviors and their relationship to different facets of user task within a classification of task types. It is a first step in our larger project to predict task type based on search behaviors.

2. RELATED WORK

2.1 User Task and Search Behaviors
Previous research has examined numerous task types with respect to their effects on user behaviors. Task types have been classified along various dimensions, standards, and definitions, including different levels of task complexity and difficulty, closed vs. open-ended tasks, known-item vs. subject search tasks, to name a few. Search behaviors have included task completion time, number of web sources used and web pages viewed, and use of web browser functions.

Task complexity has been found to impact user searching behaviors. Byström and colleagues conducted a series of studies (e.g., [4], [5]) examining the relationships between task complexity, information types, and information sources. They defined task complexity as the users' "a priori determinability of,

or uncertainty about, task outcomes, process, and information requirements" ([5], p. 194). Results indicated that task complexity was related to information type and information source selection. As task complexity increased, users needed more sources of information, more domain information and more problem solving information, were less likely to predict the types of information they needed, and were more dependent upon experts to provide useful information. Li [14] found that objective task complexity affected many aspects of searching behavior, including: the number of search systems consulted, portals visited, web result pages and items viewed, users' interaction with library resources, query-related interactive behavior, success, satisfaction, and the total task completion time. Vakkari [29] has developed a model integrating task complexity and user actions.

Many studies have classified tasks along other dimensions and examined their effects on users' search behaviors. Marchionini [19] found that users spent more time and performed more moves for an open-ended task (for which many related facts exist) than for a closed task (which had only one correct answer). Qiu [21] found that users tended to adopt more structured search patterns when engaging in specific tasks than in general tasks. Moreover, users preferred to use browsing features for completing general tasks, but in specific tasks, analytical searching was preferred. Kim [13] looked at three types of tasks: factual, interpretive, and exploratory tasks, and found that task type significantly influenced number of pages saved and the ratio of pages viewed to pages saved. Gwizdka & Spence [9] studied effects of fact-finding tasks of varying complexity on the searcher's behavior. They found that subjective task difficulty was influenced by the number of visited web pages, dwell time on a page, deviation from the optimal path, and the linearity of the navigation path. The objective task complexity was found to affect the relative importance of those factors as predictors of subjective assessment of task difficulty. Kellar, Watters, & Shepherd [11] looked at four types of tasks: fact-finding, information gathering, browsing, and transactions, and examined how users navigated and interacted with web browser across them. Results showed that information gathering task was the most complex one: participants spent more time completing it, viewed more pages, and used the web browser functions most heavily. Li [14] examined the effect of intellectual vs. decision/solution tasks on user search behaviors. Intellectual tasks were found to involve more IR systems consulted and result pages viewed, longer queries, and higher self-ratings on task success.

A variety of standards and definitions of task type classification make it difficult to compare findings across studies. Furthermore, previous studies tended to examine user behaviors at a task session level, e.g., how long a user spends on a task, instead of considering them by web page. We argue that changing focus to the web page level is useful for building user models for personalization as it may help the system learn more about the user and his task.

2.2 Dwell Time

Dwell time, or display time, is the time a user spends on a page. Morita & Shinoda [20] found that users spent more time reading documents that they rated interesting than those that they rated not-interesting. In contrast, Kelly & Belkin [12] found that using display time averaged over a group of users to predict document usefulness is not likely to work, nor is using display time for a single user without taking into account contextual factors. In particular, they found that display time differed significantly depending on specific tasks and users. One implication is that inferring document usefulness from dwell time should be tailored towards individual tasks and/or users. White & Kelly [31] further found that tailoring display time threshold based on task information improved implicit relevance feedback performance. This is evidence that display time is able to predict document usefulness when task information is considered.

Despite the seemingly conflicting findings of these studies, they all concern the relationship between dwell time and web page usefulness, or the relationship between dwell time, contextual factors (task, or user), and web page usefulness. However, these studies did not look at the relationship between dwell time and task type. It is worth examining whether search engines can learn task type from dwell time and then adapt to employ personalization for the user based on task type information.

2.3 Task and Eye Movement Behavior

Study of eye movement behavior in various visual tasks, including reading and visual search, has a long history [22]. Eye movements are cognitively controlled and visual information processing is affected by task properties in, for example, reading, face processing, scene processing and visual search ([6], [10], [16], [22], [24], [27], [29]). It is hypothesized that different visual cognition strategies are employed to meet the requirements for each type of visual task, for example the encoding of appropriate information features for the task [24]. Rayner et al. [23] found eye movement behavior for fixations and saccade distances tended to be similar for most visual tasks across individuals and cultural groups (English, Chinese, bilingual Chinese-English speakers). Reading eye movements, however, were notably different.

This suggests extended information processing interactions in service of a task may involve selection of problem solving strategies and tactics that condition parameters of the visual cognition system used to control eye movements. In this way the user's task situation could affect low-level information gathering processes. For IR the information environment is usually documents presented in a display. Study of eye movement behavior in explicit IR settings has recently received considerable attention, especially to learn details of how users process information objects (documents, web pages, etc.) [17]. Commonly used eye tracking metrics such as fixation duration, number of fixation, pupil diameter, etc., have been used as evidence of user engagement and to study patterns of eye movements associated with reading behaviors. Useful document level patterns have been identified such as the "F" shape reading pattern in a search engine result page (SERP) (e.g., [26]), and that many users read only the first few results in a SERP (e.g., [17]). Some research has compared the reading behaviors across search engines. Lorigo et al. [17] examined the number of fixations, fixation duration, and time spent on tasks for two search engines (Google and Yahoo!) and found no differences in the user processing of the results pages. Other studies have compared eye movement patterns across different IR activities.

Granka, Joachims, & Gay [7] used eye tracking data to investigate how users interact with search engine results pages using two types of search tasks, informational and navigational. They found participants viewed the first two abstracts in nearly the same way under each task and found a bias to scan the results page from top to bottom. Lorigo et al. [18] reported that task type influenced

SERP viewing time and the number of fixations on selected web documents. In informational tasks, users spent less time on SERPs and had greater pupil dilation as compared to navigational tasks.

Guan and Cutrell [8] examined whether users' search behavior was influenced when target results were displayed at various positions under navigational and informational tasks. Results indicated participants devoted more time on tasks and were less successful in finding target results when the targets were displayed at lower positions on the search results list, especially in an informational task. Eye tracking data revealed that the large decrease in performance on informational tasks for low target position ranks might be explained by the decreased probability of looking at lower positioned results.

Terai et al. [28] studied informational and transactional tasks and found participants visited more web pages for the transactional tasks, but the page reading time was shorter as compared to the informational task. Analysis of eye-movement data for 9 out of their 11 participants measured scanpath characteristics in search result pages as well as the distribution of look zones for each task, but they did not look at eye fixation data or how it may differ in different types of tasks.

In summary, previous studies of task types have found differences in users' total search time (or task completion time), number of queries, etc., for different types of tasks. However, most of these studies examined search behaviors from the task session level rather than the web page level. The majority of previous studies that examined users' dwell time on web pages focused mainly on its relationship to relevance (with or without contextual factors taken into consideration), instead of task types only. Eye tracking is acknowledged as a useful tool and Lorigo et al. [17] has underlined the need for additional research to better understand users' reading behaviors in the context of information search.

Our approach aims at identifying observable evidence search systems can exploit to learn users' task types through their behaviors so as to personalize search for the users based on task types. We examine if and how decision time, as well as users' eye movements, varies on different task types, and explore if these behavioral factors may be used as indicators of task types. We also examine several eye movement measures, total fixation number, fixation duration, average saccade distance, and the ratio of scanning to reading behaviors with respect to a detailed faceted classification of information seeking tasks. One goal is to learn if eye movement measures can be used as implicit evidence, or to confirm other evidence that is generally available, that indicates users are engaged in particular types of tasks, and so enable better support for users in achieving their information seeking goals.

3. TASKS AND THEIR CLASSIFICATION

3.1 The Task Classification

As is evident from the review of related research, how tasks have been identified and characterized varies widely (even wildly). Li's classification [15] is one of the very few examples of task classification in the information retrieval and information seeking literature which attempts to identify and integrate the various aspects, or facets, of task in a single scheme. For each of these facets she identified specific values that they could take, based on interviews with a cross-section of members of an academic community, and confirmed and extended the classification with an experimental study. The attractive feature of this classification scheme for our purposes is the ability to vary and control the values of the different facets in the construction of work and search tasks to be performed by the participants in our study. This system allowed us to relate dependent behavioral variables to a relatively small set of independent task variables, in order to identify associations.

Li [15]'s classification scheme has fifteen facets or sub-facets of work or search task. Work task is identified as the task which leads one to engage in information seeking behavior, and search tasks as the specific information seeking activities themselves. The classification itself is meant to apply to both types of tasks, and in our study, we focused on values associated with search tasks. Table 1 is an overview of the facets of Li's classification scheme which we manipulated. We added one facet, "Level", which we found to be a significant aspect of tasks in the work environment we studied. We held constant the values of the following facets (not in table 1): Source of task; Task doer; Time (length) Process; Goal (quantity); Interdependence; and Urgency, Because the values of some of Li's facets depend upon the individual searcher, the following facets are left for post-hoc analysis (not reported in this paper): Time (frequency); Salience; Difficulty; Subjective task complexity; Knowledge of task topic; Knowledge of task procedure. The choice of facets to be varied was based on Li's results, and on characteristics of typical work tasks in the journalism domain.

3.2 Our Tasks

We decided to conduct our study in the work domain of journalism, for reasons of both validity and convenience. Although journalism can be associated with any topic, it has a relatively small number of work task types. This means that we are able to have a range of topics for our tasks, while maintaining a good measure of control over realistic tasks, thus enhancing validity. At our institution we have ready access to a university journalism department, which meant both that we had experts to help us define the work tasks, and access to participants trained for such professional journalism tasks.

We began task identification by interviewing journalism faculty (including practicing journalists) about typical journalism work and searching tasks for which professional journalists receive training. The task descriptions were formalized from those interviews. We then identified a set of four of these work/search tasks which could be varied according to values of the facets which we believed could affect search behavior.

The four work tasks and associated search tasks that we identified are presented below. These tasks follow the normal scenario practice as proposed by Borlund [3], and are couched in journalism terms; that is, journalists are typically given an assignment, and an associated task to complete.

Background Information Collection (BIC): Your assignment: You are a journalist at the New York Times, working with several others on a story about "whether and how changes in US visa laws after 9/11 have reduced enrollment of international students at universities in the US". You are supposed to gather background information on the topic, specifically, to find what has already

Table 1. Facets of task which were varied in this study (After Li [15], modified)

Facets	Sub-facets	Values	Operational Definitions/Rules
Product		Physical	A task which produces a physical product
		Intellectual	A task which produces new ideas or findings
		Decision (Solution)	A task which makes a decision or solves a problem
		Factual information	A task locating facts, data, or other similar items in information systems
		Image	A task locating image(s) in information systems
		Mixed product	A task locating different types of items in information systems
Goal	Quality	Specific goal	A task with a goal that is explicit and measurable
		Amorphous goal	A task with a goal that cannot be measurable
		Combined goal	A task with both concrete and amorphous goals
Task characteristics	Objective task complexity	High complexity	A work task involving at least five activities during engaging in the task; a search task involving searching at least three types of information sources
		Moderate	A work task involving three or four activities during engaging in the task; a search task involving searching two types of information sources
		Low complexity	A work task involving one or two activities during engaging in the task; a search task involving searching one type of information source
	Level	Document	A task for which a document as a whole is judged
		Segment	A task for which a part or parts of a document are judged

been written on this topic. **Your Task**: Please find and save all the stories and related materials that have already been published in the last two years in the New York Times on this topic, and also in five other important newspapers, either US or foreign.

Interview Preparation (INT): Your assignment: Your assignment editor asks you to write a news story about "whether state budget cuts in New Jersey are affecting financial aid for college and university students. **Your Task**: Please find the names of two people with appropriate expertise that you are going to interview for this story and save just the pages or sources that describe their expertise and how to contact them.

Advance Obituary (OBI): Your assignment: Many newspapers commonly write obituaries of important people years in advance, before they die, and in this assignment, you are asked to write an advance obituary for a famous person. **Your Task**: Please collect and save all the information you will need to write an advance obituary of the artist Trevor Malcolm Weeks.

Copy Editing (CPE): Your assignment: You are a copy editor at a newspaper and you have only 20 minutes to check the accuracy of the three underlined statements in the excerpt of a piece of news story below. New South Korean President Lee Myung-bak takes office Lee Myung-bak is the 10th man to serve as South Korea's president and the first to come from a business background. He won a landslide victory in last December's election. He pledged to make economy his top priority during the campaign. Lee promised to achieve 7% annual economic growth, double the country's per capita income to US$4,000 over a decade and lift the country to one of the topic seven economies in the world. Lee, 66, also called for a stronger alliance with top ally Washington and implored North Korea to forgo its nuclear ambitions and open up to the outside world, promising a better future for the impoverished nation. Lee said he would launch massive investment and aid projects in the North to increase its per captia income to US$3,000 within a decade "once North Korea abandons its nuclear program and chooses the path to openness." **Your Task**: Please find and save an authoritative page that either confirms or disconfirms each statement.

3.3 Classification of the Four Tasks

Table 2 shows the values of the varied facets for each of the four search tasks which we gave to the participants. These values constitute the independent variables in our study, which are related to the dependent behavioral search variables.

Table 2. Variable facet values for the search tasks

Task	Product	Level	Goal (Quality)	Objective complexity
BIC	Mixed	Document	Specific	High
CPE	Factual	Segment	Specific	Low
INT	Mixed	Document	Mixed	Low
OBI	Factual	Document	Amorphous	High

BIC is a Mixed Product, because identifying "important" newspapers is intellectual, and finding documents on the topic is factual. It is at the Document Level because whole stories are judged; it has the Specific Goal of finding documents on a well-defined topic; it has High Objective Complexity because of the number of sources and activities that need to be consulted/done.

CPE is a Factual Product, because facts have to be identified; it is at the Segment Level, because items within a document need to be found; it has the Specific Goal of confirming facts; it has Low Objective Complexity because only three facts need to be confirmed.

INT is a Mixed Product, because defining expertise is intellectual, and contact information is a fact; it is at the Document Level, because expertise is determined by a whole page; Goal Quality is Mixed, because determining expertise is amorphous but contact information is specific; it has Low Objective Complexity because only two people need to be found.

OBI is a Factual Product, because facts about the person are needed; it is at the Document Level because entire documents need to be examined; Goal Quality is Amorphous because "all the information" is undefined; it has High Objective Complexity because many facts need to be found.

4. METHODOLOGY

4.1 Study Design

A lab-based study was designed to investigate the effects of search task type and task facets (described in the previous section) on searching behavior, such as saving and reading behaviors. Data was collected on a variety of searcher behaviors, such as eye gaze, and various interactions with the search systems and information objects, with the goal of relating various of these behaviors to explicit statements of tasks and task facets.

4.2 System

The experiment described in this paper was designed and conducted using a system that reduces the complexity of creating interactive information retrieval (IIR) experiments that log users' multidimensional interactive search behavior [2]. The system has a client-server architecture where researchers configure IIR experiments from a range of extensible tasks. The current experiment configuration applied assigned tasks, questionnaires to gather background information and perceptions before and after the tasks, and usefulness evaluation questionnaires. The system was used to rotate tasks into sequences and monitor the progress of the experiment. Users accessed the experiment through an interface that presented them with their task sequence and provided them with additional instructions. The system is able to log a wide range of user behaviors with an array of heterogeneous logging tools. For the experiment, logs were created for web traffic using UsaProxy (http://fnuked.de/usaproxy/) and Morae (http://www.techsmith.com), keyboard and mouse activity using RUI (http://ritter.ist.psu.edu/projects/RUI/), and eye movements using the Tobii T60 eyetracker with Tobii Studio (http://www.tobii.com). The experiment system framework is available as open source (http://sourceforge.net/projects/piirexs/).

The search interface in the experiment system has two frames: on the right side is the regular Internet Explorer (IE) window, with a blank starting page; on the left side is a panel that allows the users to save desired pages and also to delete them. Figure 1 depicts the search interface with two saved web pages.

Figure 1. The Search Interface

4.3 Participants

We used a convenience sampling method, recruiting students from the undergraduate Journalism/Media Studies program in our School to mimic journalists. To ensure that the participants have a certain writing skills, only upper-division undergraduates who had completed either one journalism writing or reporting class were selected. For this study, we had planned to recruit a total number of 32 participants, and till the data analysis for the current paper, 22 (18 female, 4 male) had finished the experiments. Participants were recruited from relevant writing and reporting classes personally, with distributed flyers, and via targeted emails. They were informed in advance that their payment for participation in the experiment would be $20.00, and that the 8 who have saved the best set of pages for all four tasks, as judged by an external expert, would receive an additional $20.00.The rationale for the extra payment was trying to ensure that participants treat their assigned tasks seriously. The participants were between 18 and 27 years old. Most students spoke English natively (73%) with the remainder of the population stating a high degree of English knowledge. Participants rated their computing skills high with an average search experience of 8.5 years using a range of different browsers (IE 32%, Firefox 64%, as well as others). Students rated their search experience generally high but claimed more experience with WWW search as compared to online library catalog search. They were generally positive about their average success during online search.

4.4 Procedure

Each participant was given a tutorial as a warm-up task and then performed four web search tasks (described in section 3). Although the experiment setting was controlled, participants were free to go anywhere on the Web using IE 6.0 to search for information and were asked to continue the search until they had gathered enough information to accomplish the task. During the search, all of the participants' interactions with the computer system were logged. Their gaze during the search was recorded using the eye-tracking system. The entire search process was stored via the Morae screen-capture program. In each task, when participants decided they found and saved enough information objects for purposes of the task, they were then asked to evaluate the usefulness of the information objects they saved, or saved and then deleted, through replaying the search using the screen capture program. An online questionnaire was then administered to ask about their searching experience, including their subjective evaluation of their performance, and reasons for that evaluation. The order of the four tasks was systematically rotated for each participant following a Latin Square design for a total of 32 participants. After completing four different tasks, an exit questionnaire was administered, asking about subjects' perceptions of their search experiences, the extent to which they found differences in the tasks, their ability to perform the tasks, and their overall search experiences in the tasks.

4.5 Eye Movement Data

The eye movement data was collected using a Tobii T-60 eye tracking display, which logs eye gaze position at 60 Hz. The display resolution was 1280x1024. We used the eye fixation data as calculated by the Tobii algorithm. The eye tracker was calibrated for each participant before the tutorial task and collected data covering all of the tasks in the experiment. The logs were processed to extract the eye fixation data for analysis.

Most of the previous eye tracking work in IR settings has reported reading behavior that could be more accurately described as aggregates of eye gaze position ('hot spots') without distinguishing the fixation subsequences that comprise true reading behavior. As part of our analysis of the eye fixation data we created an algorithm that implements the E-Z Reader reading model [25] to identify reading fixation sequences as distinguished from isolated fixations, which we define as 'scanning' fixations. Scanning fixations also provide information to a person, although the amount is limited to that available in the foveal (in focus) field. A collection of fixations in a reading sequence provide more information, both because information is gained from the larger parafoveal region, and, of course, because of the richer semantic structure available in sentences, etc. as compared to isolated units of several words. Importantly, some of the types of semantic information available through reading sequences may be critical for a user to satisfy task requirements. Beymer & Russell [1] provide a good description of the issues in processing gaze data to extract reading sequences.

The reading model has been used to measure a number of parameters of reading behavior and classify fixation data to create reading state transition models. It can also be used to provide more robust investigation of actions during document dwell time that may shed light on relationships to relevance and task effects, such as those reported by Kelly & Belkin [12]. A basic application of the reading model is to classify eye fixations as reading or scanning to investigate the effect of tasks on the ratio of reading to scanning fixations. We report those results in the present work.

4.6 Behavioral Measures

The independent variables used in the analysis and presented in this paper include the task and task facets described in section 3. The dependent variables include the following:

- Task completion time: the total time users spent to complete a task;
- Number of web pages visited in a task;
- Number of queries issued in as task;
- Number of sources: number of unique Internet domains visited by a participant in a task;
- Number of search sources: number of unique search engines or databases used by a participant in a task;
- Decision time: time taken during the search process to decide whether a document is useful;
- Reading to scanning fixations ratio: the ratio of the number of reading to scanning fixation (described in Section 4.5);
- Average saccade distance: the distance between the positions of two successive eye-fixations calculated in screen coordinates.

5. RESULTS

5.1 Overall Search Behavior by Task

Table 3 presents the overall task level behaviors including task completion time, number of web pages visited, number of queries, number of information sources, and number of search sources they have used in each task by all the participants. The task completion time of BIC and CPE were not normally distributed, and the number of search sources of all four tasks was not normally distributed, so for these two behavioral variables, the non-parametric Kruskal-Wallis H test was used. All other variables: number of pages visited, number of sources, and number of queries, were normally distributed in all four tasks, and one-way ANOVA was used for analyzing them.

Significant differences were found for all 5 measures across the four tasks. Post-hoc analyses using Tukey's test found that users spent significantly longer time to accomplish BIC than the other three tasks, while there is no difference in time in the other three. Users visited significantly less pages in CPE than in the other three tasks, and less pages in INT than in BIC. They used significantly more sources in BIC and OBI than in INT and CPE. Number of queries had the similar pattern: they issued significantly more queries in BIC and OBI than in INT and CPE. In terms of search sources, users went to more sources in the BIC and OBI tasks than in CPE, and more sources in BIC than in INT.

Table 3: Overall search behaviors in each task

Overall behavior	Mean (Standard deviation)				Test statistics
	BIC	CPE	INT	OBI	
Task completion time (min.)	17.79 (6.42)	8.03 (6.04)	10.80 (6.44)	12.65 (5.78)	$\chi^2(3,N=88)=22.56$, **p<.001**
# of pages visited	47.77 (18.91)	15.41 (9.85)	28.82 (14.86)	39.77 (18.41)	$F(3,84)=17.09$, **p<.001**
# of sources	16.91 (6.94)	7.05 (3.51)	11.05 (4.72)	15.95 (6.73)	$F(3,84)=14.44$, **p<.001**
# of queries	17.41 (10.19)	5.86 (4.68)	9.36 (5.49)	15.55 (9.54)	$F(3,84)=10.27$, **p<.001**
# of search sources	3.14 (1.89)	1.50 (.86)	1.64 (.79)	2.73 (1.93)	$\chi^2(3,N=88)=21.75$, **p<.001**

Note: # denotes "number".

5.2 Overall Search Behavior by Task Facet

5.2.1 Products of Search Tasks

Along the task Product facet, the distributions of all five measures in the two groups were not normal, so the non-parametric Mann-Whitney U test was used to test the differences for all measures. It was found (Table 4) that users spent significantly longer time to complete Mixed-Product tasks than Factual tasks. They visited significantly more pages and more sources in Mixed-Product tasks than in Factual tasks. However, the number of queries they issued did not show differences, nor did the number of search sources they used.

5.2.2 Objective Task Complexity

For the Objective Task Complexity facet, the distributions of average number of queries in the two groups were normal, allowing a t-test. The other four measures in the two groups were not normal, so Mann-Whitney U test was used. Results (Table 4) show that users spent significantly longer time to complete High Complexity tasks than Low Complexity tasks. They visited significantly more pages and more sources, issued significantly more queries, and used significantly more search sources in High Complexity tasks than Low Complexity tasks.

5.2.3 Task Level

For the Level facet, the non-parametric Mann-Whitney U test was used to test the differences because the distributions of all five measures in the two groups were not normal. Results (Table 4)

Table 4: Overall search behaviors by task facets

Overall behavior	Task Product		Task Complexity		Level		Task Goal (Quality)		
	Factual (CPE,OBI)	**Mixed (BIC, INT)**	**Low (CPE,INT)**	**High (BIC,OBI)**	**Document (BIC,INT,OBI)**	**Segment (CPE)**	**Specific (CPE,BIC)**	**Mix (INT)**	**Amorphous (OBI)**
Task completion time (min.)	10.34 (6.29)	14.29 (7.28)	9.42 (6.33)	15.22 (6.57)	13.75 (6.81)	8.03 (6.04)	12.91 (7.89)	10.80 (6.44)	12.65 (5.78)
	U(86)=663, z=-2.55, p<.05		**U(86)=494, z=-3.96, p<.001**		**U(86)=376, z=-3.37, p=.001**		χ^2(2, N=88)= 1.43, p=.489		
# of pages visited	27.59 (19.10)	38.30 (19.35)	22.11 (14.19)	43.77 (18.88)	38.79 (18.91)	15.41 (9.85)	31.59 (22.14)	28.82 (14.86)	39.77 (18.41)
	U(86)=656.5, z=-2.60, p<.01		**U(86)=342.5, z=-5.22, p<.001**		**U(86)=194,, z=-5.13, p<.001**		χ^2(2, N=88) = 4.56, p=.102		
# of sources	11.50 (6.96)	13.98 (6.57)	9.05 (4.59)	16.43 (6.77)	14.64 (6.64)	7.05 (3.51)	11.98 (7.38)	11.05 (4.72)	15.95 (6.73)
	U(86)=730.5, z=-.199, p<.05		**U(86)=356, z=-5.12, p<.001**		**U(86)=215, z=-4.93, p<.001**		**χ^2(2, N=88)= 6.98, p=.03**		
# of queries	10.70 (8.89)	13.39 (9.05)	7.61 (5.35)	16.48 (9.80)	14.11 (9.20)	5.86 (4.68)	11.64 (9.77)	9.36 (5.49)	15.55 (9.54)
	U(86)=757, z=-1.76, p=.078		**t(86)=-5.27, p<.001**		**U(86)=283, z=-4.28, p<.001**		χ^2(2, N=88) = 5.13, p=.077		
# of search sources	2.11 (1.60)	2.39 (1.61)	1.57 (.82)	2.93 (1.90)	2.50 (1.72)	1.5 (.86)	2.32 (1.67)	1.64 (.79)	2.73 (1.93)
	U(86)=793, z=-1.53, p=.125		**U(86)=471, z=-4.36, p<.001**		**U(86)=424, z=-3.06, p<.001**		χ^2(2, N=88) = 4.36, p=.113		

Notes: 1. # denotes "number". 2. The values reported in this table are means (standard deviations) unless specified.

show that users spent longer time to complete Document Level tasks than for Segment Level tasks. They visited more pages and more sources, issued more queries, and used more search sources in Document Level tasks than in Segment Level tasks.

5.2.4 Task Goal (Quality)

For the Goal (Quality) facet, the distributions of all five measures in the two groups were not normal, and the Kruskal-Wallis H test was used to test the differences. It was found (Table 4) that only average number of sources showed differences in the different categories. The post-hoc analysis using Tamhane found that users visited more sources in Amorphous tasks than in Mixed tasks ($p<.05$). The other measures had no significant differences.

5.3 Decision Time by Task and Task Facet

Decision time analysis was focused on content pages only, excluding querying pages and search result pages, since it makes more sense to discuss on the usefulness of content pages than on others. We also removed some pages that were not actually read by the users and were served only for navigational purposes. For example, in the experiment, for logging purpose, users were asked not to open a new window or close the window while searching, so they often used the 'BACK' button in the browser to return to the search result list pages and/or other previous pages. In such cases, the pages that had to be displayed before the users reached their target pages were navigational pages. Based on the observation in the experiment and the examination of the dwell time distribution, we decided to use 1.8 seconds as the threshold to identify the navigational pages, and so web pages with a dwell time of less than 1.8 seconds were removed for the analysis.

5.3.1 By Tasks

Results show that the mean decision time (in seconds) for the four tasks were 19.46 (BIC), 28.95 (CPE), 18.73 (INT) and 16.01 (OBI). The Kruskal-Wallis H test found significant differences among the four tasks ($\chi^2(3, N=1499) = 58.52$, $p=.000$). The post-hoc Tamhane analysis show that the mean decision time for CPE was significantly longer than for the other tasks ($p<.001$). The other three tasks did not have statistical differences.

5.3.2 Products of Search Tasks

The Mann-Whitney U test revealed no differences for decision time among tasks with different products of search tasks (Table 5). The product of the search tasks did not seem to influence users' decision time for each web page.

5.3.3 Search Level

The Mann-Whitney U test revealed significant differences for decision time between Segment level and Document level tasks (Table 5). The decision time for the Segment level task was significantly longer than that for Document level tasks.

5.3.4 Goal (Quality)

The Kruskal-Wallis H analysis revealed significant differences for decision time among tasks with different search goal quality (Table 5). The more specific the search goal, the longer the decision time was. Post-hoc analysis using Tamhane found that users had significantly longer decision time for the Specific tasks than for the Amorphous task and the Mixed task.

Table 5: Decision time for different facet values

Facets	values	Decision time in secs. (SD)	Test statistics
products of search tasks	Factual (CPE,OBI)	19.89(18.28)	U(1497)=256465, z =-1.86 p = .06.
	Mixed (BIC,INT)	19.15(18.74)	
Level	Document (BIC,INT,OBI)	18.12(17.26)	U(1497)=81695, z=-7.14, **p<.001**
	Segment (CPE)	28.95(24.00)	
Search goal (quality)	Amorphous (OBI)	16.01(13.50)	χ^2(2, N=1499) =29.33, **p<.001**
	Mixed (INT)	18.73(20.18)	
	Specific (CPE,BIC)	22.02(19.94)	

5.4 Task Effects on Eye Movements

The eye movement results are limited to 20 subjects due to technical difficulties extracting data for two subjects. Of the eye movement statistics considered, only average saccade distance data was not normally distributed. The average saccade distance and reading to scanning ratio exhibited significant task effects.

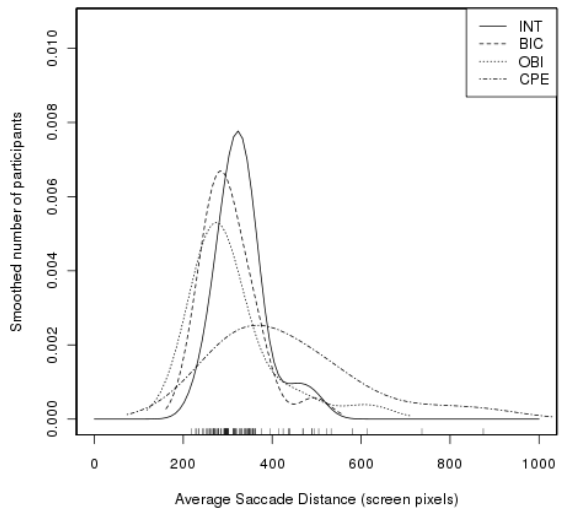

Figure 2. Average saccade distance by task

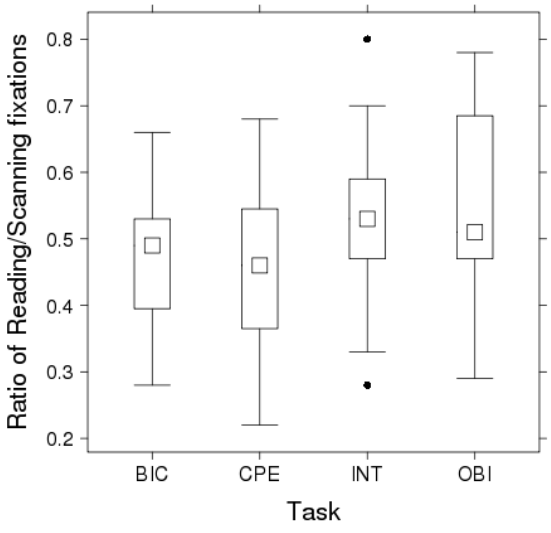

Figure 3. Ratio of reading to scanning behaviors by task

Friedman's nonparametric test found the tasks had a significant influence on the average saccade distance ($\chi 2$ = 19.23 (df=3), p=0.0002). Figure 2 shows a single task, CPE, is primarily responsible. Using ANOVA confirms a significant effect for tasks on average saccade distance (F(3,20)=6.62, p<.001, r=0.58). Tukey multiple comparisons of means confirmed that CPE is significantly distinguished from the others tasks by average saccade distance.

For the reading to scanning ratio data tasks were also found to have a significant effect, F(3,20)=3.09, p<.05, r=0.58. Tukey multiple comparisons of means shows, however, that only OBI and CPE are significantly distinguished (Figure 3).

5.5 Task Facet Effects on Eye Movement

Two task facets, task complexity and document engagement level, were found to have significant effects on eye movement statistics. Task complexity was found to have a significant effect on average saccade distance (Figure 4), F(1,20)=9.41, p<.001. The level of document engagement in a task, document or segment, was also found to have a significant effect on average saccade distance, F(1,20)=15.52, p<.001.

Level had a significant effect on the reading to scanning fixation ratio, F(1,20)=5.83, p<.05, r=1. CPE was significantly different from OBI, BIC and INT from this perspective (Figure 5).

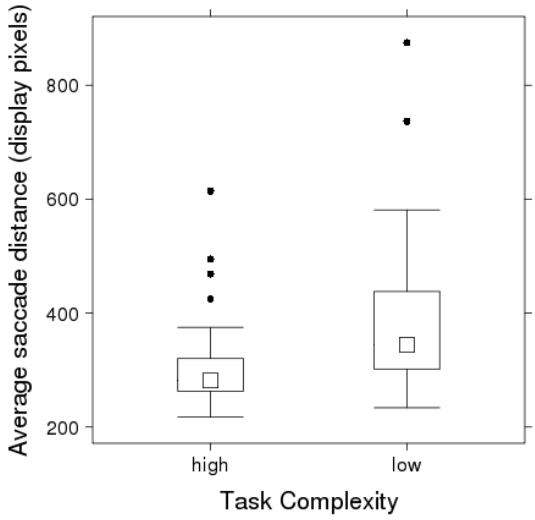

Figure 4. Effect of complexity on average saccade distance

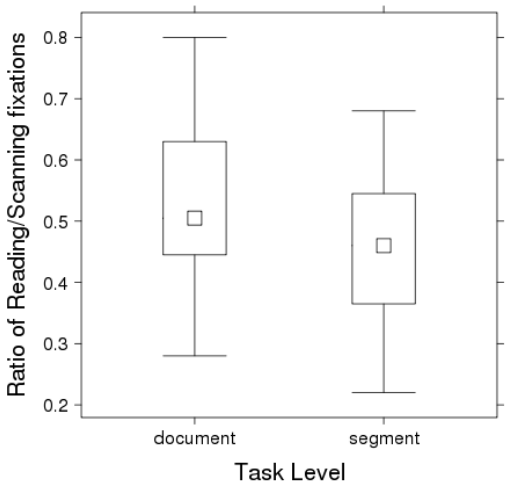

Figure 5. Reading to scanning ratio by task level

6. DISCUSSION

The experiment was designed to examine task behavior using realistic tasks in an unconstrained setting, so the creation of faceted tasks was a particular challenge. The tasks and several of the task facets were distinguishable by various measurements we made, including total completion time, decision time, number of sources used, and the average saccade distance and reading to scanning fixation ratio. Here, we comment on two aspects of our results: the validity of our tasks with respect to objective task complexity, and interpretation of differences in behaviors with respect to the cognitive and other factors associated with the different facet values.

The results for decision time showed we were successful in controlling for complexity. The participants interacted with significantly more information sources and unique URLs in high complexity tasks than in low complexity tasks. It is interesting to see confirmation of task complexity effects on these high level cognitive behaviors in the saccade distance and reading to scanning ratio eye movement behaviors.

The CPE task is especially interesting because it was distinguished in several behaviors by several task facets, uniquely by level. Decision time in CPE was significantly longer than in all other tasks. This can be well explained by the facet Level. The other three tasks were on Document level, for which users only needed to make judgment on the usefulness of the whole documents. CPE was a Segment level task, for which users had to look for specific pieces of information in a document in addition to locate and judge its general usefulness. The segment level information search requires users to engage pages in a more careful manner, which tended to prolong the time spent to make a usefulness decision.

The average saccade distance for copy editing was significantly longer than for the other tasks, and the reading to scanning ratio was biased towards scanning. Copy editing was unique among tasks in the facet "Level" where segments of documents were to be judged rather than entire documents. It seems plausible participants might adopt an information processing strategy of scanning a document for particular facts or snippets rather than reading for detail. The distinction between reading and scanning fixations may be important in this regard. It has long been known a skipping sequence of fixations with retracements is characteristic of reading. Given the explicit goal of finding and evaluating facts in the copy editing task, it seems plausible users engaged in this type of task would adopt an information search strategy of scanning for words representing individual facts. Implementing this strategy would, on balance, result in longer saccades. In contrast, an effective information search strategy for search tasks requiring evaluation of entire documents might be expected to include more reading and shorter average saccade distances. So the effect of the CPE task on average saccade distance and reading to scanning ratio may be explained by the task facet Level.

Table 6 shows the significant associations of behaviors with different facet values. Although, taken alone, they do not provide enough information to make decisions about the association of a particular behavior with a task facet, they do provide a potential basis for doing so. Predictive power is likely to require additional information, such as threshold values, or identification of conjunctions of observed behaviors. Note that Level was associated with all of the behaviors, and since CPE was the only example of a difference in this facet one has to be careful in interpreting those associations. Task is also significantly associated with all of the behaviors, but it is CPE that is distinguished from all of the other tasks, so the association of task against behaviors appears to be due to the uniqueness of the copy editing task.

Table 6. Summary of significant results

Observable search behavior	Task	Task facets			
		Comple-xity	Product	Level	Goal (quality)
Task completion time	X	X	X	X	
# of pages visited	X	X	X	X	
# of sources	X	X	X	X	X
# of queries	X	X		X	
# of search sources	X	X		X	
Decision time	X			X	X
Read/scan ratio	X			X	X
Saccade distance	X	X		X	

7. CONCLUSIONS

The results reported in this paper, although certainly not conclusive with respect to our overall goal of being able to predict task facet values (that is, task *type*) from search behaviors, are a important step in that direction. Significant associations were found between a large variety of behaviors and task facet values, indicating that it should be possible to identify decision points for predicting these task facet values. We have not yet analyzed interaction effects amongst the different facet values. That analysis could help in identifying specific decision rules. Post hoc analysis of the user-oriented facet effects, at least some of which could be identified in operational environments through the searchers' past behaviors, may also support the decision task. These analyses are the next step in our overall project.

As always, there are significant limits to our study as well as need for future efforts. The small number of participants is a clear problem. The study will be completed with a total of 32 participants, which, although not a great number, will help. The study used a convenience sampling method, recruiting journalism/media studies students as mimicking professional journalists. A more serious issue is that the design is unbalanced by facet values, with only one task having a specific facet value. In addition, post-hoc analysis using information collected from questionnaires is not analyzed and reported in this paper.

Despite these limitations and the need for further efforts, the results as they stand are promising. We ran an experiment using realistic tasks related directly to the backgrounds of the participants, who searched in the live Web, with no constraints on what they could do there, and found significant differences in behaviors in this quite unconstrained environment. It seems realistic to believe that we will eventually be able to predict task type from user behaviors, as they take place, and to use task type to help interpret other implicit sources of evidence relevant to effective personalization of the search experience.

8. ACKNOWLEDGMENTS

This research was sponsored by IMLS grant LG 06-07-0105.

9. REFERENCES

[1] Beymer, D., & Russell, D. M. (2005) WebGazeAnalyzer: A system for capturing and analyzing web reading behavior using eye gaze, CHI '05, 1913–1916.

[2] Bierig, R., Cole, M., Gwizdka, J., Belkin, N.J., Liu, J., Liu, C., Zhang, J., & Zhang, X. (2010). An experiment and analysis system framework for the evaluation of contextual relationships. In Proc. of the 2nd Workshop on Contextual Information Access, Seeking and Retrieval Evaluation, Milton Keynes, UK, March 28, 2010.

[3] Borlund, P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. Information Research, 8(3), paper no. 152. Retrieved from http://informationr.net/ir/8-3/paper152.html

[4] Byström, K. (2002). Information and information sources in tasks of varying complexity. Journal of the American Society for Information Science and Technology, 53(7), 581-591.

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

[6] Findlay, J. M., & Gilchrist, I. D. (2003). Active vision: The psychology of looking and seeing. Oxford University Press.

[7] Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In SIGIR '04, 478–479.

[8] Guan Z. & Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. In CHI '07, 417–420.

[9] Gwizdka, J., Spence, I. (2006). What Can Searching Behavior Tell Us About the Difficulty of Information Tasks? A Study of Web Navigation. Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), vol. 43. Austin, Texas, USA.

[10] Henderson, J. M. (2003). Human gaze control during real-world scene perception. Trends in Cognitive Science, 7, 498–504.

[11] Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing Web-based information-seeking tasks. Journal of the American Society for Information Science & technology, 58(7), 999-1018.

[12] Kelly, D., & Belkin, N.J. (2004). Display time as implicit feedback: Understanding task effects. SIGIR '04, 377-384.

[13] Kim, J. (2006). Task as a predictable indicator of information seeking behavior on the Web. Unpublished dissertation, Rutgers University.

[14] Li, Y. (2008). Relationships among work tasks, search tasks, and interactive information searching behavior. Unpublished dissertation. Rutgers University.

[15] Li, Y. (2009) Exploring the relationships between work task and search task in information search. Journal of the American Society for Information Science and Technology, 60, 275-291.

[16] Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. Trends in Cognitive Science, 4, 6–14.

[17] Lorigo, L., Haridasan, M., Brynjardottir, H., Xia, L., Joachims, T., Gay, G., Granka, L.,Pellacini, F., & Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. Journal of the American Society for Information Science & technology, 59(7), 1041-1052.

[18] Lorigo, L., Pan, B., Joachims, T., Granka, L. & Gay, G, (2006). The influence of task and gender on search and evaluation behavior using Google. Information Processing & Management, 42(4):1123–1132, 2006.

[19] Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. Journal of the American Society for Information Science, 40(1), 54-66.

[20] Morita, M., & Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In SIGIR '94, 272-281.

[21] Qiu, L. (1993). Analytical searching vs. browsing in hypertext information retrieval systems. Canadian Journal of Information and Library Science, 18(4), 1-13.

[22] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research, Psychological Bulletin, 124, 372–422.

[23] Rayner, K., Li, X., Williams, C.C, Cave, K. R., & Well, A.D. (2007) Eye movements during information processing tasks: Individual differences and cultural effects. Vision Research, 47(21), 2714-2726.

[24] Rayner, K., Smith, T.J., Malcom, G.L., & Henderson, J.M. (2009) Eye movements and visual encoding during scene perception. Psychological Science 20(1):6-10.

[25] Reichle, E., Pollatsek, A, & Rayner, K. (2006) E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading, Cognitive Systems Research 7(1), 4–22, 2006.

[26] Sherman, C. (2005). A new F-word for Google search results. Search Engine Watch.

[27] Starr, M.S., & Rayner, K. (2001). Eye movements during reading: Some current controversies. Trends in Cognitive Science, 5, 156–163.

[28] Terai, H., Saito, H., Egusa, Y., Takaku, M., Miwa, M., & Kando, N. (2008). Differences between informational and transactional tasks in information seeking on the Web, in Proceedings of IIiX 2008.

[29] Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychological Review, 113, 766–786.

[30] 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.

[31] White, R., & Kelly, D. (2006). A study of the effects of personalization and task information on implicit feedback performance. In CIKM '06, 297-306.

[32] White, R.W., Ruthven, I., & Jose, J.M. (2005). A study of factors affecting the utility of implicit relevance feedback. In SIGIR '05, 35-42.