Task and user effects on reading patterns in information search

Michael J. Cole, Jacek Gwizdka *, Chang Liu, Ralf Bierig, Nicholas J. Belkin, Xiangmin Zhang

School of Communication & Information, Rutgers, The State University of New Jersey, 4 Huntington Street, New Brunswick, NJ 08901, USA

ARTICLE INFO

ABSTRACT

We report on an investigation into people's behaviors on information search tasks, specifically the relation between eye movement patterns and task characteristics. We conducted two independent user studies ($n = 32$ and $n = 40$), one with journalism tasks and the other with genomics tasks. The tasks were constructed to represent information needs of these two different users groups and to vary in several dimensions according to a task classification scheme. For each participant we classified eye gaze data to construct models of their reading patterns. The reading models were analyzed with respect to the effect of task types and Web page types on reading eye movement patterns. We report on relationships between tasks and individual reading behaviors at the task and page level. Specifically we show that transitions between scanning and reading behavior in eye movement patterns and the amount of text processed may be an implicit indicator of the current task type facets. This may be useful in building user and task models that can be useful in personalization of information systems and so address design demands driven by increasingly complex user actions with information systems. One of the contributions of this research is a new methodology to model information search behavior and investigate information acquisition and cognitive processing in interactive information tasks.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Extended interaction with an information system to obtain and utilize information is a prime example of a complex cognitive task. Understanding the cognitive aspects of information seeking to improve user interfaces and better 'fit' information systems to individual users to improve their efficiency and effectiveness is an important direction for research and development of information search systems (Belkin, 2008). Important contributors to the complexity of interaction with information search systems are the variable nature of the user's task, the range of displayed content that must be accessed and presented, and the cognitive characteristics of the individual and their cognitive stance, for example their information acquisition strategy, during the interaction. Reading is essential for acquiring information during primarily textual information search. Importantly, eye movement is cognitively controlled (Findlay and Gilchrist, 2003) and actively used in service of user tasks (Triesch et al., 2003). Reading eye movement behavior has been studied for many years and a good deal is known about the details of its role in text information acquisition (Rayner, 1998). Models of reading eye movement patterns using eye gaze location and the duration of fixations, i.e. repeated eye gaze, have been developed (Reichle et al., 2006). To our knowledge these

models have not been applied to analyze interactive information retrieval over task sessions.

In this work, we use reading models to explore several dimensions of this complex interaction environment. Since reading eye movement behavior is essential to acquiring (textual) information in a search session, it is reasonable to ask how characteristics of an individual's reading model are related to factors known to influence information search behaviors, such as the task type, level of user task knowledge, and individual cognitive abilities. These relationships are important to investigate for fundamental reasons, given the essential role of reading in search information acquisition. Analysis of search interaction using reading models can also have practical implications. Calculation of reading models requires only sequential observation of the user's eye fixation location and duration and can be constructed on the fly. If we can learn relationships between reading model parameters and factors that are not available for observation in most situations for search systems, such as individual differences and task type, then we may be able to predict aspects of the user and their information environment and adapt the system to the user.

A specific goal of our research project is to build personalized systems that exploit the task characteristics to improve the effectiveness of interactive search and information retrieval (IR). One way to build better adaptive systems is to construct richer user models based on inferences about the cognitive state of the user. Eye tracking information can inform user models and enhance performance by including cognitive features related to user attention and intent (Conati et al., 2005). Connecting aspects of the

user's situations, e.g. task characteristics, with unobtrusive measurements of user cognitive strategies is a specific way to adapt information systems to the user's goal. Acquisition of eye data for use by information retrieval systems has been investigated to detect relevance (Buscher et al., 2008a; Oliveira et al., 2009) and provide real-time relevance feedback for query expansion (Buscher et al., 2008b).

Research to model and analyze cognitive interactions of whole navigation and search sessions has been conducted (Juvina et al., 2005). Our efforts are distinguished from this work in the focus on building models of task session interactions at the web page interaction level. Our work investigates eye movement reading patterns and their relationships to individual differences, web page types, and search task types. In this paper we present evidence for significant individual differences in reading models, some task effects, and effects of different Web page types visited during information search.

2. Background and related work

2.1. Task effects on search behavior

One aspect of user context that affects information search behavior is the nature of the user's search task. The effects of task characteristics, including complexity, difficulty, and stage, on search behavior, usefulness and relevance judgments, have been studied extensively (White and Kelly, 2006; Li, 2009; Gwizdka, 2008; Byström and Järvelin, 1995; Liu and Belkin, 2010). Kelly and Belkin (2004) found that document usefulness could be better predicted with implicit measures such as dwell time on retrieved document when the user's task was considered. They concluded document display time is not likely to be an accurate indication of document usefulness when display time is averaged over a group of users. Even at the level of an individual user, it is important to take account of contextual factors to achieve reasonable performance in predicting document usefulness from display time. Attempts to use dwell time to infer a document's usefulness to a task therefore stand the best chance of success when one has knowledge about the individual and their task. This study did not address which contextual factors should be included or how such factors might be incorporated into a system to predict document usefulness from dwell time. Also not addressed was the potential improvement in information system performance that might be achieved by implementing this approach. White and Kelly (2006) made some progress in addressing the potential effectiveness of taking account of contextual factors by looking at task and user influences on dwell time. Their study exploited the idea of having a system that displayed a document for an optimal amount of time based on information about the task and user. They manipulated the document display time to limit the potential dwell time on the document as a function of task and user characteristics and were able to establish a dwell time threshold to predict document usefulness in the context of user and task. The results showed the display time threshold based on task information could improve performance of implicit relevance feedback algorithm. This shows display time is predictive of document usefulness when task information is considered.

2.1.1. Task stage and task types
Consideration of different granularity of units of interaction in information seeking tasks has received research attention. This work has explored how these more global properties of task sessions influence user search behavior and performance. It has been suggested that information seekers have different levels of information need as they progress through task stages or phases (e.g.

Kuhlthau, 1991; Lin, 2001; Taylor, 1968). Cognitive and affective user states and the actions they take vary with these different stages. This research indicates task stage may be an important factor that relates to the user's judgment of document usefulness. Vakkari and colleagues (e.g., Vakkari and Hakala, 2000; Vakkari et al., 2003) found that users' search tactics change along stages, for example, their users were, at the beginning of their tasks, less likely to start their initial queries by introducing all the search terms, were more likely to enter only a fraction of the terms, and tended to use more synonyms and parallel terms than in later stages. They also found that users' relevance criteria depend on the stage of their task performance process, although their results did not show a statistical significance of the changes. Taylor et al. (2007) found statistically significant relationships between users' relevance criteria and task stage choices corresponding to those of selection, exploration, formulation, and presentation in Kuhlthau's ISP model (Kuhlthau, 1991). This branch of research demonstrates the differences in users' relevance judgments in different stages of the task; however, it leaves as an open issue how such differences could be modeled through the user's behaviors. In addition, the above described research considered only the effect of task stage but did not correlate it with user behaviors to suggest how to implicitly build user models and improve system performance to better help people search. More research is needed to study how task stage can be helpful in adapting search to different users.

Accurate identification of task stages is difficult because clear temporal boundaries are not manifest. One can do a priori gross divisions by time; say into thirds of the session length (White et al., 2005). Another approach to investigate task stage effects is to cast the tasks to be completed into several sub-tasks. Lin (2001) manipulated the user task, by giving distinct sub-tasks to be completed in different search sessions. The task scenario in his study required participants to make a vacation plan through three steps in three sessions: identifying candidate places, choosing one place, and making a trip plan. Both approaches to operationalize task stage are somewhat arbitrary. The ability to make task and sub-task distinctions in user studies is valuable because analysis of task behavior at sub-task and finer levels of interaction sequences appear to offer better generalizability of research results because it may well be easier to identify these task units in complex information task behaviors.

Task type classification has received considerable attention (Freund, 2008; Kellar et al., 2007; Li and Belkin, 2008). Li and Belkin (2008) have proposed a comprehensive scheme to classify tasks based on dimensions of task features. We use this task classification scheme in our present work.

2.1.2. Level of knowledge
Domain knowledge or expertise has been shown to have significant effects on task performance and search behavior (e.g. Hsieh-Yee, 1993; Kelly and Cool, 2002; Toms et al., 2007; Wen et al., 2006; White et al., 2009; Zhang et al., 2005; Wildemuth, 2004). Knowledge impacts search tactics and strategies. For example, Wildemuth (2004) showed that lower domain knowledge had negative effects on the efficiency of query tactics and increased query reformulation errors. Hembrooke et al. (2005) reported that domain knowledge was positively correlated with the effectiveness of search strategies.

Prior knowledge has been correlated with search performance measures. Duggan and Payne (2008) administered a knowledge test before participants searched using the questions in the test. They found the search performance score was positively correlated with the participant's knowledge score. Acquisition of knowledge during search has also received attention. Lawless et al. (2007) investigated the relationship between prior knowledge and the

acquisition of new knowledge during web browsing by treating one group of participants with reading selected to enhance their domain knowledge prior to the search. Pre- and post-test knowledge scores of the treated and untreated groups showed that participants in the treatment group significantly increased their knowledge score.

2.1.3. Cognitive ability and individual differences: influences on search and navigation behavior

Juvina and van Oostendorp (2006) looked at individual cognitive differences and user behaviors to build an empirical model of web navigation. They found that an individual's spatial and semantic cognitive abilities were important determinants in successful execution of web navigation tasks.

Ford et al. (2005) conducted a user study (n = 68) and classified 507 user queries from three web search tasks of increasing complexity. Participants were classified along the lines of approaches to studying, various cognitive and demographic properties, and their understanding of Web-based information seeking and their typical approach to gaining information in Web-based searching. They found consistent patterns between increasing task complexity and combinations of individual differences.

2.1.4. Eye tracking use to investigate web search behaviors

Eye tracking metrics such as fixation duration, number of fixations, and pupil diameter, have been used as evidence of user engagement and to study patterns of eye movements associated with reading behaviors. Farzan and Brusilovsky (2009) used eye tracking to explore the usefulness of social navigation clues to users performing web searches. Document level patterns have been identified, e.g. an "F" shape reading pattern for a search engine result page (SERP) (Sherman, 2005) and that many users read only the first few results in a SERP (Pan et al., 2007).

Some research has compared reading behaviors across search engines. Granka et al. (2004) used eye tracking data to investigate how users interact with SERPs during two types of search tasks, informational and navigational. They found the mean fixation duration for the two top ranked items was nearly the same, and that there was a top to bottom scanning bias in a SERP. Lorigo et al. (2008) examined the number of fixations, fixation duration, and time spent on tasks for Yahoo! and Google and found no differences in the user processing of SERPs. Task type was shown to influence 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 (2007) examined task type influences on user search behavior by manipulating the positions of target results in navigational and informational tasks. Overall, participants devoted more time to tasks and were less successful in finding target results when the targets were displayed at lower positions on the search results list. This effect was especially pronounced for informational tasks (as opposed to navigational tasks). The eye tracking data showed that there was a decreased probability of looking at results in the low position, explaining the poor performance. Terai et al. (2008) studied informational and transactional tasks. They found that participants visited more web pages when performing the transactional task, but with shorter page dwell time as compared to the informational task. Eye-movement scan paths were described for individual participants when looking at search result pages, and a distribution of 'look zones' for each task were identified. This work did not provide an analysis of the influence of the task types on the eye fixation measurements.

2.2. Eye movements and visual cognition strategies

Eye fixations provide information about attentional states because the only way to acquire information visually is by repeated eye gaze on a location. Research has established that eye movements are cognitively controlled (Findlay and Gilchrist, 2003) and are conducted in service of user tasks (Hayhoe et al., 2007). One way to discover user intent during task performance is to couple observable behaviors with inferences based on fixation observations. For example, fixations tell us which words have been lexically-processed, allowing one to connect semantics with explanations for an interaction behavior, such as query reformulation or document selection and use.

Visual information processing is affected by immediate task properties in reading, face processing, scene processing and visual search (Reichle et al., 1998; Rayner et al., 2009; Findlay and Gilchrist, 2003; Torralba et al., 2006). It is hypothesized that different visual cognition strategies are employed to meet the requirements for each type of task such as the encoding of appropriate information features for the task (Rayner et al., 2009). There is some understanding of the probable mechanism by which these types of task differences cause changes in fixation patterns (Hayhoe et al., 2007). Eye movement behavior for fixations and saccade distances has been found to be similar for low-level visual tasks, for example face recognition and visual search, across individuals and cultural groups. Reading behavior is notable in that individual and cultural differences have been observed (Rayner et al., 2007). This suggests that extended sequences of 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 task could affect low-level information gathering processes.

Visual information acquisition is important for information seeking, for example in reading to process texts. Taking account of task effects on visual cognition, for example via effects on reading eye movement control, may therefore be a valuable component in user models to personalize search and information retrieval in adaptive systems.

People can be classified by their patterns of eye movements in processing texts. Hyönä et al. (2002) investigated individual reading patterns in expository text using eye tracking. The participants were asked to read a text passage and produce a summary, which was assessed for accuracy. Fixation patterns were recorded for six types of sentences and the sequences of sentence processing, marking revisits to a previous sentence or look ahead fixations to an upcoming sentence. Cluster analysis identified four groups of individual patterns, which were characterized as reading strategies. The groups were distinguished by whether they refixed on previous sentences, engaged in look-ahead fixations to upcoming sentences, and whether they concentrated on structural features in the text presentation, for example, processing headlines and subject headings. The working memory of the individuals was assessed using a reading span test. Individuals with the largest working memory tended to use the strategy of focusing on the structural features and were also the best performers in the experiment.

Hyönä et al. (2002) provides evidence for individual differences in eye movement reading patterns for processing information, and for classification of individuals by reading patterns. Their work also suggests these patterns can be explained in part by appeal to use of cognitive strategies in extracting information from texts. They also provide evidence for connections between reading patterns and individual cognitive abilities, at least with respect to working memory.

The designed structure of pages have effects on eye movement processing of text and cognitive evaluation of the text content. For example, SERP pages with entries laid out in grids vs. lists induced different eye movement patterns to process the entries and resulted in different evaluations of the trustworthiness of the items (Kammerer and Gerjets, 2010).

The work of Kammerer and Gerjets (2010) suggests web pages with different layout structure, such as SERPs and content pages may be processed differently by users. Analysis by page type may throw light on individual differences and task type effects during text-based search interactions.

2.3. The EZ Reader reading model

The EZ Reader model is a cognitively-controlled, serial-attention model of reading eye movements (Rayner et al., 1998; Pollatsek et al., 1999, 2003, 2006; Reichle et al., 2004, 2006). It takes word identification, visual processing, attention, and control of the oculo-motor system as joint determinants of eye movement in the reading process. The EZ Reader model is a processing model and is based on the assumption that reading is a cognitively-controlled process where the saccade to the next word is programmed while the person is cognitively processing the text available in the currently attended fixation. The saccade programming has a labile stage. If the next word is recognized during this labile stage, the programmed saccade is canceled and a saccade to the next word is programed.

There is a distinction between high acuity text in the foveal region and progressively lower acuity text in the parafoveal region. It has been shown that some text information can be extracted from the parafoveal region and frequently it allows a saccade to jump over several words. The role of textual information available in the parafoveal region has been studied extensively, (e.g. Blanchard et al., 1989; Inhoff and Rayner, 1986; Kambe, 2004; Liu et al., 2002; Morris et al., 1990; Pollatsek et al., 1986; Rayner et al., 2003; White and Liversedge, 2005). There is evidence for recognition and use of word length (Juhasz et al., 2008), orthographic features, and some semantics such as the predictability of the word in context (Drieghe et al., 2007) and morphological features (Drieghe et al., 2010).

Processing of text during fixations has been shown to take place in several stages. Orthographic recognition takes place in about 40 ms, followed shortly thereafter by phonological recognition ($\sim$60 ms), and then lexical processing, which takes at least 113 ms (Reingold and Rayner, 2006). There is a labile lexical processing period, from 113 ms to 168 ms, during which the next saccade programming can be reprogrammed with a new saccade target. After that labile period, the pending saccade will be executed after the cognitive processing is completed. This is one way in which observations of eye movements can be connected with the semantics of information processing. Eye movements are cognitively controlled and the eyes remain fixated during the lexical processing period independently of the stimuli, for example even if the word is removed (Findlay and Gilchrist, 2003). The next saccade takes place after lexical processing is completed. It has long been known that familiarity and conceptual complexity of the text processed is positively correlated with the fixation duration (e.g. Rayner and Duffy, 1986).

The basic EZ Reader model does not account for higher-order cognitive processes, for example those involving language comprehension and conceptual processing. While this is a limitation of the model, it is claimed the model provides an explanation of the moment-to-moment reading process when linguistic processing is running smoothly (Reichle et al., 2004). The EZ Reader model is well-suited to our work, which is directed at correlating measures of information acquisition during information search with high-level search behaviors rather than the details of semantic cognitive processing.

3. Objectives and research questions

Reading behavior in interactive information search task sessions is of interest because it is necessary for acquiring information. Reading to acquire information is a linguistic and cognitive process that might be able to explain the information-related aspects of information seeking tasks.

As the literature review shows, research in information science and HCI has established that task, and individual differences, such as knowledge and cognitive abilities, affect search behaviors. It makes sense then, to look at models of the low-level reading process in task sessions and explore the influences of task and individual differences on the observed interaction reading models. The layout and nature of content in differing web pages is another potential influence on reading models of interactive search sessions.

Information acquisition through eye fixations comes mostly in the form of reading, i.e. making use of lexical knowledge that recognized glyphs are meaningful words, coupled with linguistic knowledge to understand sentences and larger units of meaning. This is the critical path by which a person connects textual information acquisition with their knowledge. A reading model of textual information acquisition is therefore a natural approach to representing one aspect of how humans carry out information tasks.

Do users adopt information acquisition strategies that are influenced by their task, knowledge, and/or cognitive capabilities? Effects of these factors on user behaviors have been demonstrated in patterns of queries (Wildemuth, 2004; Liu et al., 2010). In this work, we explore relationships between a model of reading behaviors and task, knowledge, and a measure of working memory ability.

For information search in a predominantly textual environment, the parameters of reading model and eye movements can be hypothesized to result from individual characteristics, the properties of the user's information task, and properties of the displayed information (e.g. a web page) processed by the user.

$$ReadingModel_{parameter} = f(individual, task_{type}, page_{type})$$

This provides a framework to address distinct dimensions of the general problem of information search in textual environments.

Our work concerns reading during realistic information seeking tasks. The goal is to explore the relationship between selected reading model parameters and eye movements and the above factors in order to learn if reading models can be employed to predict user characteristics, properties of the page being processed, or the user's task type during a search session. We want to know the degree to which each component is correlated with observable eye movement pattern properties and whether interaction effects exist. This work concentrates on relationships with reading eye movement pattern representations, especially reading models.

The relationship between the individual's properties, search tasks and web page types and reading eye movement parameters can be analyzed at several levels. We will begin by looking at task effects on reading model parameters and then move to a more detailed view of task interactions by looking at effects due to web page types.

This work uses the data collected in user studies designed to examine task type and individual differences (such as, cognitive abilities in one study, and domain knowledge in another) effects on information search session behaviors. The studies employed realistic search tasks designed to differ along specific dimensions.

We explore two questions associated with the reading model influence hypothesis:

(1) Does task type affect reading model parameters?
(2) Do the reading model parameters change when processing different types of textual information content, specifically, different types of web pages?

4. Methodology

In this paper we report on results from two user studies, one using information search tasks in the journalism domain and the other using genomics domain tasks. The general goal of our research is to learn how to identify search behaviors that can be used to model users and personalize interactions with information systems over extended task sessions.[1] The journalism study considered the effect of different types of tasks on search behaviors, and the genomics study looked at the search behavior effects of an important individual property, differences in domain knowledge.

Both experiments used a simulated work task approach (Borlund, 2003) to design the tasks as presented to the participants in order to enhance the generalizability of the results. While the experiment setting was controlled, this bias towards realistic tasks in an extended interactive setting means many variables were not controlled. For example, in the journalism study participants were free to go anywhere on the web to search for information to meet the task goal. This choice in study design makes for significant limitations in analyzing results. On the other hand, a strength of the studies is that the eye movement observations were of natural search behavior. The genomics study included a sophisticated search engine with both a simple and advanced search interface for a domain-specific search application. Participants were free to choose between them during the task session. In the journalism study, the viewed content was unrestricted and subject to typical variations expected in operational environments with web pages of various formats and types, including mixtures of text and images, some changing dynamically with advertisements.

Each experiment was conducted using the multi-source logging POoDLE system (Bierig et al., 2009). During search in both experiments, all of the participants' interactions with the computer system, including eye gaze, were logged. Eye data was collected with a Tobii T-60 eye-tracker (1280 × 1024 @ 60 Hz).[2] We used eye fixation data as calculated by the Tobii Studio software (foveal radius = 35 pixels).

4.1. Experiment 1: journalism domain tasks

This user study investigated behaviors associated with different task types for 32 advanced undergraduate journalism students carrying out realistic professional journalism tasks. Each participant was given a tutorial and performed four tasks involving web search (described below). Participants were asked to continue searching until they had gathered enough information to accomplish the task or 20 min had elapsed. Logging issues for two participants prevented complete data analysis. In the following we report on results for 30 of the 32 participants.

4.1.1. Tasks and task classification

The journalism work domain is attractive for investigation of task type effects on searching behavior because there are a small number of task types, but they can be associated with any topic. This means topics can be varied in 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 and were able to get expert help to define the work tasks along with access to participants trained for such professional journalism tasks.

A set of four tasks was identified by interviewing journalism faculty and practicing journalists (Appendix A). Li (2009) provides

a task classification system which attempts to identify and integrate the various aspects, or facets, of task in a single scheme. For each facet specific values they can take have been identified based on interviews with a cross-section of members of an academic community. The classification scheme has been confirmed and extended with an experimental study. This classification system allows us to describe the work tasks and identify categorical differences between them. The tasks were designed to vary according to values of the characteristics which we believed could affect search behavior.

Li'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. We added a "Level" facet to Li's classification scheme because we found it to be a significant aspect of tasks in the work environment we studied. We held constant the values of other facets, including: Source of task; Task doer; Time (length) Process; Goal (quantity); Interdependence; and Urgency. The choice of facets to be varied was based on Li's results, and on characteristics of typical work tasks in the journalism domain.

4.1.2. Tasks

Tasks were generated by first 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 afforded some variation in the values of the facets which we believed could affect search behavior.

The four work tasks used were: advanced obituary (OBI), interview preparation (INT), copy editing (CPE), and background information (BIC). The task descriptions are provided in Appendix A. The task format follows the scenario practice as proposed by Borlund (2003) and they were presented to participants in a realistic manner, by giving the journalism students an assignment and an associated task to complete.

4.1.3. Classification of the four tasks

Table 1 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. The tasks varied in several dimensions: complexity defined as the number of necessary steps needed to achieve the task goal (for example, identifying an expert and then finding their contact information), the task product (factual vs. intellectual, e.g., fact checking vs. production of a document), the information object (a complete document vs. a document segment), and the nature of the task goal (specific vs. amorphous).

BIC task 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.

Table 1
Journalism study task characteristics (task product: F = factual; I = intellectual).

Task	Product	Level	Goal	Complexity
OBI	F,I	Document	Specific	High
CPE	F	Segment	Specific	Low
INT	F,I	Document	Mixed	Low
BIC	F,I	Document	Amorphous	High

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. In Table 1, one can see the advanced obituary (OBI) and the copy editing (CPE) tasks have the least similarity.

4.1.4. Experiment design and participants

A laboratory-based study was designed to explore 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 task and task facets.

For this study, 32 upper-division undergraduate journalism students were recruited from our University. They were informed that their pay for participation in the experiment, either $20.00 or $40.00, would depend on performance as judged by experts. The rationale for the extra payment was to try to insure 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 reporting 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 on-line library catalog search. They were generally positive about their average success during on-line search. All participants were required to have completed either one journalism writing or one reporting class.

4.1.5. Procedure

Each participant filled out a demographic questionnaire and then performed two tests of cognitive abilities. One measured operation span, which is related to working memory capacity. The other test measured mental rotation ability, which may relate to field closure ability. A tutorial on the system using a warm-up task followed. They then performed four web search tasks (described in Section 3). Before beginning each search a questionnaire was administered to learn about their task understanding and anticipated difficulty. Although the experiment setting was controlled, participants were free to go anywhere on the web using IE (v6) 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 interactions with the computer system, including eye gaze, were logged. The entire search process was stored via the Morae[3] screen-capture program. When participants decided they found and saved enough information objects for purposes of the task, they were then asked to evaluate the useful-

Table 2
Task (TREC topic) characteristics.

TREC task id	Difficulty	Specificity	Category
2	Hard	Specific	Genetic structure
7	Easy	General	Genetic processes
42	Very easy	Specific	Genetic phenomena
45	Hard	General	Genetic phenomena
49	Easy	General	Genetic structure

ness of the information objects they saved, or saved and then deleted, while replaying the search using the screen capture program. An on-line questionnaire was then administered to ask about their searching experience, including their subjective evaluation of their performance, and reasons for that evaluation.

After a training task, participants completed the four tasks in counterbalanced order. 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.

4.2. Experiment 2: genomics domain search tasks

The purpose of this exploratory experiment was to look for effects of differences in domain knowledge on search behaviors. The search tasks and expert document relevance judgments were taken from the 2004 TREC genomics track (Hersh et al., 2005). The available tasks were categorized along two dimensions: hard vs. easy and general/specific where specificity refers to path length from the most general node to the task topic subject in the MeSH tree.[4] Hard topics were those with few relevant documents, as judged by the expert TREC assessors, returned by our search system using the task description as the query. We used the number of relevant documents returned in the top ten results as the performance measure. Each topic had a topical MeSH category that reflected its location the MeSH ontology trees.

Five tasks were used and the task questions were the TREC genomics track descriptions. These tasks were designed to be examples of information tasks for research professionals. These types of search tasks are difficult even for medical librarians (Liu and Wacholder, 2008). The tasks used in the study are listed in Appendix A.

A simulated work task approach (Borlund, 2003) was used to design the task presentation to the participants. Participants were asked to find and save all of the documents useful for answering the task questions.

Table 2 lists the tasks selected for the study and their characteristics. We switched task 49 for task 42 during the study because task 42 was very easy, so the results for these tasks involved different participants.

4.2.1. Search collection

We implemented a search system using Indri from the Lemur toolkit.[5] The search collection was taken from the TREC genomics collection, a 10-year, 4.5 million document subset of the Medline bibliographic database (Hersh et al., 2005). We used the documents from the 2000–2004 period ($n = 1.85$ million) to allow for reasonable retrieval performance. The TREC ad hoc retrieval tasks were based on 50 topics relating to five general types (Roberts et al., 2009).

4.2.2. Study participants

---

[3] http://www.techsmith.com/morae.asp.

[4] http://www.nlm.nih.gov/mesh/mtr_abt.html.
[5] http://www.lemurproject.org/.

Forty students from the authors' institution participated in the study. They were recruited from related schools and departments, including biology, pharmacy, animal science, biochemistry, and so on. Undergraduate students, graduate students, and post-docs participated in the study. The number of graduate students and the number of undergraduate students were roughly balanced to elicit the different levels of knowledge. Post-docs were assigned to the graduates group.

Technical difficulties prevented analysis of eye gaze data for two participants. Of the 38 participants, 15 were non-native English speakers and 23 were native English speakers. Another two of the non-native English speakers are removed from the analysis for the present work because they exhibited reading difficulties.

4.2.3. Domain knowledge measurement

At the beginning of the experiment participants rated MeSH indexing terms in the three MeSH trees associated with the task topic categories: genetic processes (G05), genetic phenomena (G13), and genetic structures (G14). They were asked to rate their knowledge on each term using a five point scale: 1 – "no knowledge", 2 – "vague idea", 3 – "some knowledge" 4 – "high knowledge", 5 – "can explain to others". The participant domain knowledge (PDK) was calculated by simply summing the participant's self-assessed knowledge of each rated term and normalizing by the rating an expert would give. That is:

$$PDK = \frac{\sum wp_i * t_i}{5 * m}$$

where wp_i is the individual knowledge rating for the MeSH term and i ranges over the rated MeSH terms. m is the total number of terms to be rated ($m = 409$ in the study) and t_i is 1 or 0, depending on whether the term was rated by the participant. The hypothetical expert is presumed to have rated all terms as 'can explain to others' (i.e. 5). While the PDK values can range from 0.0 to 1.0, notice that most will range between 0.2 and 1.0, because a term rating of 1 means 'no knowledge'. If a participant rated every term in a topic as 'no knowledge' their PDK would be 0.2. A participant could score below 0.2 only if they failed to rate terms in the topic. Of course a PDK could be greater than 0.2 and also include unrated terms.

4.2.4. Experiment procedure

The subjects read and signed a consent form and filled out a questionnaire about their background, computer experience and previous searching experience. They were then asked to rate their knowledge of MeSH terms in three categories related to the search tasks (total – 409 terms). Before each task the subjects filled out a questionnaire that elicited self-reports of their familiarity with the task and level of knowledge along with a judgment of the expected difficulty of the task. They were then allowed up to 15 min to conduct the search task. All of the tasks were recall-oriented, that is the participants were asked to find and save as many useful documents as they could. After ending the search task, the participants evaluated the relevance of each saved document. Finally, they filled out a questionnaire asking them to assess the actual task difficulty and the amount of learning they experienced. The interaction with the system was logged by the computer. After completing all tasks they filled out an exit questionnaire that elicited comments about their impression of the realism of their search experience with their real-world experience of similar tasks. The experiment was conducted in a human–computer interaction lab and each participant was tested individually and received $25.

4.3. Reading models

In most previous eye tracking work in information search settings reports of reading behavior have been based on analysis of eye gaze position aggregates ('hot spots'), without distinguishing the fixation subsequences that comprise true reading behavior. The eye fixation analysis in this work is based on an implementation of the EZ Reader reading model (Reichle et al., 2006).

4.3.1. Implementation of the EZ Reader reading model

Our implementation of the EZ Reader model is line-oriented. That is, we do not address the problem of smooth transitions in reading lines in a block of text. Reading of two consecutive lines in a text is coded as two instances of reading sequences. We also do not distinguish the type of reading content in applying the algorithm. A section heading is treated the same as a line of text in a paragraph. This is a limitation of the current implementation of the algorithm. In this work we do not analyze the results in terms of larger text structures such as paragraphs, or by text structure elements, such as headings. Our model extends the basic EZ Reader by allowing for regression fixations to previous positions in the line of text processed. Such regressions have been long known to be a common feature in reading and have been correlated with the processing of unfamiliar words or sentences with greater structural or conceptual complexity (Rayner and Pollatsek, 1989).

The inputs to our algorithm are successive fixation locations and their duration. The output is a classification of the sequences of fixations as members of a reading sequence, or as isolated lexical fixations, which we call 'scanning' fixations. Reading sequences and scanning instances are restricted to lexical fixations, that is, fixations that exceed the lexical processing threshold of 113 ms. The foveal (in focus) region is operationalized by taking the Tobii default of 35 pixels for projection of the foveal radius on the display. The display projection of the parafoveal annulus was taken to extend from 35 pixels to 120 pixels. As suggested by the EZ Reader model, we distinguish between the left hand side of the parafoveal region and the right hand side to model the left to right reading pattern of the English text presented in the studies. Our algorithm should generalize to many languages with similar orthographic encoding, i.e. non-logogram or phonetically-encoded glyphs. Perceptual span describes the region of text that can be taken in, on average, with a single fixation. It has been shown to be a function of the orthographic system, with readers of logographic systems, for example Chinese, exhibiting a perceptual span of three or so characters (Inhoff and Liu, 1998) as compared to 8–9 characters typical for readers of English or Dutch (Pollatsek et al., 1986).

The basic idea in the algorithm is that reading proceeds by a sequences of fixations where the succeeding fixation is within the region previously looked at. This region may be stepwise extended to the right, provided the new fixation is within the ride-hand-side parafoveal region. Fixation regressions to any point in the previously fixated text are allowed. It is possible to extend the text region on the left hand side by regression to the beginning of the reading area. Saccade landing positions are subject to Gaussian errors around the target. Research has shown that although the center of the fixation may be above or below the text centerline, people read just the line of text and not words above or below (Hornof and Halverson, 2002).

The process identifies which fixations are part of a reading sequence. A reading sequence ends when a fixation is not in the right-hand-side parafoveal region and is not a regression to a point in the line of text already swept out in the reading instance.

The algorithm was used to distinguish reading fixation sequences from isolated fixations, which we define as 'scanning' fixations. Scanning fixations provide some semantic information, limited to that available in the foveal visual field (Rayner and Fischer, 1996). Fixations in a reading sequence provide more information both because information is gained from the larger parafoveal region (Rayner et al., 2003), and because of the richer semantic structure available in compositions of text, including sentences,

paragraphs, etc., as compared to isolated units of several words. Importantly, some of the types of semantic information available through reading sequences may be crucial for satisfying of user task requirements. Our reading model is a fixation classifier that identifies fixations in eye movement sequences that indicate processing of text that is more than an isolated word. Our use of 'scan' and scanning must be understood in this sense. Reading behavior that might typically be described as 'scanning', that is a disciplined processing of text by skipping some of the text and reading phrases, would be modeled as a series of short reading sequences by our model.

The reading model algorithm was used to classify fixations as reading or scanning. This classification of fixations was used to create reading state transition models (see an example in Fig. 1) for each participant and task pair.

A user's decision to read or scan is a distinctive cognitive commitment and our analysis focuses on the Scan to Read, and the Read to Scan transition probabilities. These decisions reflect the visual information acquisition strategy adopted by the user to achieve their task goal. It is useful to notice that this hypothesizes a two state reading model that reflects user intent with respect to acquiring textual information in the course of their task. More complicated state models can be constructed using our reading model methodology by identifying criteria to classify patterns of reading and/or scanning fixation sequences. Criteria for classification could be strictly sequential, for example identifying a pattern of a sequence of short reading sequences, or could involve other dimensions, such as structural elements in text, say, section titles, or regions on a page.

There is a practical attraction of our approach to characterizing task information acquisition by modeling reading fixations. The reading model construction only requires analysis of the recent eye movement sequence to classify observed fixations. One consequence is the methodology allows for construction of reading models on the fly and so enables real time detection of parameter changes. There is potential for use of reading model parameter values in operational settings to implicitly gain information about the user and their situation and to personalize the system interactions.

4.4. Detailed task-session analysis

The layout of text on a page influences patterns of eye movements, for example as shown by Kammerer and Gerjets (2010). The layouts of pages viewed by participants were not controlled in our studies. In the journalism experiment they varied in the typical ways a web page might vary. In the genomics experiment, the formatting was uniform, but the abstract content could vary in length and style reflecting the practices of a variety of journals. It is intuitively plausible that users process search engine results pages (SERPs) and content pages differently because their immediate information task is different. In the former case, the dominant task is to identify a promising document link and in the latter, to acquire information that is specifically useful to the user's main task. Therefore, analysis of results by page type may throw light on individual differences and task type effects during text-based search interactions.
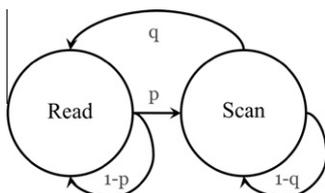

Fig. 1. A general diagram of a two-state *reading model*.

This paper includes an investigation using reading models to gain insight into task effects at the page level. Reading models were calculated from the eye fixations that took place on each page. Various additional eye movement measures, such as total number and duration of lexical fixations, were also calculated. In addition, we characterized some details of reading sequences within a page, including the length (in pixels) of the text read, the average saccade distance within each reading sequence, and the number of retrograde saccades within reading sequences executed by a user on a page.

The eye-movement data in task sessions for the journalism experiment was further analyzed by considering the user interaction in different types of interaction units that we defined by the type of web page. We distinguished two basic classes: SERPs and the content pages reached by a user following a link in a SERP. This was something of a convenience as the design of the experiment allowed participants free access to the world wide web. Yet, it is also true that these page types have distinctive roles in interactive information retrieval. SERPs are typically processed with the goal of recognizing links to potentially fruitful documents. The goal in processing content pages is to determine if it contains information that can be somehow useful to the user's overall task, so information acquisition from the page to determine usefulness and extract pertinent information is the main concern.

In addition to analyze the SERP and content page types, we decided to look at highly visually attended pages. This was operationalized by selecting a subset of the available data where there was data available for all four tasks. We selected pages on which the total lexical fixation duration on the page was above the median for all pages in the task for the participant. This criterion helps to eliminate pages from analysis that contained more than a couple of eye tracking errors. This procedure may introduce a bias into the data we analyzed because below median fixation duration pages may be of high quality and reveal significant aspects of the interaction with the information system. For example, participants may have dismissed a page quickly in deciding on its relevance to the task. Development of more sophisticated criteria to weed out low quality pages will receive attention in our future work using the reading model methodology. It is worth noting that some bad eye tracking data are not errors. For example, participant eye blinks will result in missing data, as will cases where the participant looks off-display. In both of our experiments the task descriptions were available on another display during task performance.

The analysis of the data from the journalism experiment was restricted to 17 participants for the content page analysis and 18 for the SERPs. This introduces a minor distortion in comparing content and SERP results when the measures were sums over all participants, as for example in the total text acquisition. We choose to have these slightly different sets in order to maximize the observations in this exploratory study. The normalized measures and state transition probabilities can be directly compared, although the content page data will have slightly more noise.

5. Results

Recall that one goal of our overall research project to personalize interactive information retrieval is to predict the type of task which leads a person to engage in information search, based on the person's behaviors (in our case, eye movements) during the course of the search. We therefore have analyzed our data from this point and view, and present the results of our analyses according to their response to our research questions:

(1) Do parameters of a person's reading model change with task type?

(2) Do the reading model parameters change when processing different types of textual information content, specifically, different types of web pages?

5.1. Task Influence on an individual's Reading model

Changes from a user's typical reading model due to a task depends on some weighting of the influence of the task characteristics, and consequently the task demands on the user for successful task completion. The task characteristics in our two classification systems are categorical so for an initial analysis it is reasonable to suppose their influence is a weighted linear combination. Ranking tasks by differences in the changes in the state transition probabilities compared to each participant's typical reading model provides a description of the task influence on the model. Not having knowledge of the participant's typical model, we calculate an average reading model for each participant by summing the transition probabilities observed for the four tasks and taking the mean. The impact of each task was then calculated as the difference in the observed transition value from the average model mean. The participant's tasks were then ranked for each state transition in the model. For each state transition the set of task ranks was collected for all participants and Friedman's ANOVA rank sum test was applied to determine if there was a correlation between the tasks and the impact on individual reading models. Table 3 reports the values obtained for the data from experiment 1 and shows tasks affected both the Scan to Read and the Read to Scan state transitions.

Fig. 2 plots the variance in the Scan to Read and Read to Scan transition probabilities from the average transition probability over all of the tasks for each participant for the journalism tasks. Most participants were biased by OBI (advanced obituary task; task level: document) to more frequently switch from scanning to reading. For CPE (fact checking task; task level: segment) most participants were biased towards scanning.

The Read to Scan transition occurs when the participant arrives at the end of a reading sequence and could transition to another reading sequence or switch to scanning (Fig. 1). In those cases, the opposite behavior was observed. For OBI, the bias to switch to scanning was suppressed and they tended to continue to another reading sequence. For CPE, participants were biased to switch to scanning. INT appears to have had an effect similar to OBI, and BIC had little impact.

Turning to the genomics search tasks, we did not find task effects on Scan to Read or *Read to Scan* transition probability. This reflects the relative uniformity of genomics tasks with respect to task facets.

5.2. Task and page type effects

5.2.1. Task effects on text acquisition

A direct, but rough, measure of textual information acquisition by reading is to count the number of pixels in a horizontal line covered by the lexical eye fixations. One output of our reading model implementation is the pixel length of the reading sequences identified in the input sequence of fixations. The total length covered in
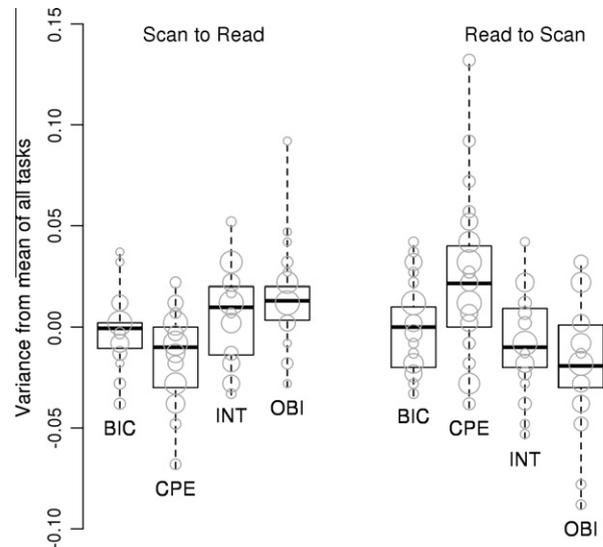

Fig. 2. Journalism task effects on participant reading state transition probabilities.

isolated lexical fixations, i.e. 'scan' fixations, is calculated simply as 70 pixels per scan fixation based on Tobii's default 35 pixel foveal radius mapping to the display. Our methodology allows for an estimation of both the total display length of lexical processing and distinguishing isolated text acquisition and the amount acquired in reading groups of words.

The total text acquired over the course of task, and the means by which it is acquired may reveal differences between task types. To investigate this, we calculated the pixel length of the reading subsequences in each page using the reading model algorithm. We report these text acquisition results as number of display pixels in the plots. This is not easily interpreted however, so from time to time in presenting the results we will convert these measurements to the number of words acquired. To do this, we sampled a number of content and search pages in the journalism experiment as recorded in the Tobii movie and calculated a mean of 0.016 glyphs per pixel. The average word length in the English language is 4.5 (Shannon, 1951), and so the rough conversion to number of words acquired is: $numWords = (readingPixelLength * 0.016)/4.5$.

Fig. 3 (left hand side) shows the total text acquisition, which is the sum of the text acquired in reading sequences and from the isolated lexical fixations, for content pages and for SERPs. For content pages the total text acquired was greatest for BIC and least for CPE, but the difference was not quite significant. However, total text acquisition in SERPs by task was significant. For total text acquisition by scanning BIC and OBI were highest and CPE and INT were lowest (Kruskal–Wallis chi-squared = 15.31, $p < 0.002$). The same was true for total text acquired in reading sequences on SERPs (Kruskal–Wallis chi-squared = 21.80, $p < 0.0001$). Likewise for the total text acquisition where the same pattern was observed (Kruskal–Wallis chi-squared = 23.51, $p < 0.0001$). Roughly, the mean number of words acquired in each task was: BIC (263), CPE (97), INT (91), and OBI (184).

These results may be seen to reflect a facet of the journalism task categorization (Fig. 3-left). BIC and OBI are distinguished from CPE and INT by high complexity. Recall that in Li's task classification scheme complexity is the number of steps required to complete the task.

When the average amount of text acquired is considered (Fig. 2 right hand panel) there is no significant difference between the tasks on SERPs using the Kruskal–Wallis non-parametric test. For content pages, there is a significant task difference for CPE (Kruskal–Wallis chi-squared = 7.86, $p < 0.05$).
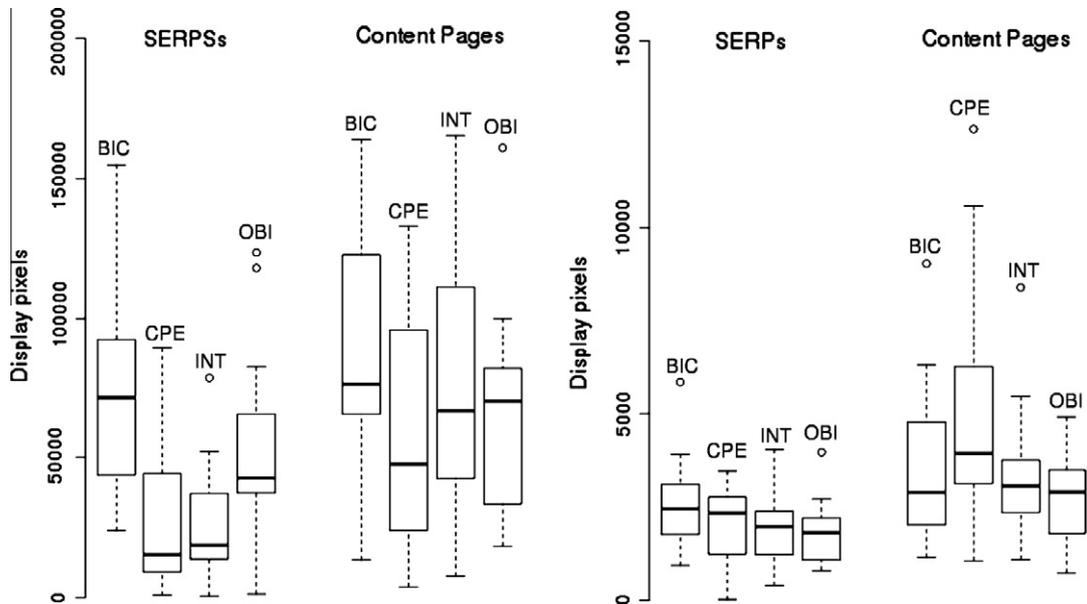
Table 3
Friedman's rank sum test for task effects on participant reading model parameters.

Reading model state transition	Friedman χ^2	df	p-Value
Scan → Read	17.7	3	<0.001
Read → Read	3.3	3	0.35
Read → Scan	17.0	3	<0.001
Scan → Scan	2.9	3	0.41

Fig. 3. Total text acquired (left panel) and per page average text acquired (right panel) in screen pixels for the two page types (SERP and content).

The existence of task influences on text acquisition in total text acquired and text acquired on average pages makes it worthwhile to further analyze task influences by looking at effects on individual pages. To do this we looked at the state transitions on pages and on the contributions of reading sequences and isolated lexical scanning fixations to the information acquisition during a task.

5.2.1.1. State transitions and page types. We counted the number of *Scan to Read* and *Read to Scan* state transitions and calculated the number of these transitions on a typical SERP and a typical content page. These average transition counts provide some insight into the engagement and use of SERP page types and content page types.

In Fig. 4 it is clear that the BIC task induces the largest number of both Scan to Read (Kruskal–Wallis chi-squared = 12.18, $p < 0.01$) and Read to Scan (Kruskal–Wallis chi-squared = 13.25, $p < 0.005$) transitions on an average SERP. Again, BIC's combined task goal and complexity facets distinguish it from the other tasks. It is plausible that the amorphous nature of the task goal may induce more switching on SERP pages to identify potentially useful links.

State transitions in content pages tell a different story. Fig. 4 (right hand panel) shows the situation for state transitions on an 'average' content page. Here the CPE task asserts itself. The distinctly higher number of both Scan to Read (Kruskal–Wallis chi-squared = 9.43, $p < 0.05$) and Read to Scan (Kruskal–Wallis chi-squared = 9.57, $p < 0.05$) state transitions reflects the task level facet. The copy-editing (CPE) task is to be addressed by small units of text, specifically a fact in a context, so it is to expected that someone performing such a task might adopt a strategy of looking for a key word (or number in this case) and then read to confirm the context. All of the other tasks had a "document" value of the task level facet, that is, processing of the entire page to determine its usefulness for the task.

5.2.1.2. Text acquisition by reading vs. scanning. The reading model algorithm enables distinguishing the amount of text acquired in reading sequences from the text acquired in isolated lexical fixations. Recall that, apart from the ability to gain some insight into the spatial processing of the page, there is a qualitative difference in the semantic content that can be acquired in these uses of reading. That is, scanning enables only extraction of information from

individual words or short phrases, while reading in extended fixation sequences can acquire more sophisticated meaning from words, for example by processing sentences. So it is not unreasonable to think that tasks with different characteristics might bias information acquisition strategies in the mixture of reading by isolated scanning and reading sentences, etc. Further it is reasonable to think that SERPs and content pages may reflect differences in text acquisition given their distinct roles in the interactive information retrieval process.

Fig. 5 (left hand panel) shows that average text acquisition by scanning on SERPs and content pages. The right hand panel shows the average text acquisition via reading sequences. In text acquisition in reading sequences, CPE had more text acquired on an average content page (Kruskal–Wallis chi-squared = 7.86, $p < 0.05$) as compared to the other tasks. The same was true of text acquisition in scan fixations on content pages (Kruskal–Wallis chi-squared =13.14, $p < 0.005$).

Turning to text acquisition from reading in isolated scan fixations (Fig. 5 left hand panel), we found a similar pattern in task influences when compared to the text acquisition via reading sequences. These effects were significant in SERPs (Kruskal–Wallis chi-squared = 15.31, $p < 0.001$) and in content pages (Kruskal–Wallis chi-squared = 7.86, $p < 0.05$). On SERPs, BIC is again significantly different and supports the observations on SERP page state transitions. It is reasonable to hypothesize that overall, participants tended to allocate more effort to processing SERPs when the task goal was amorphous as compared to other tasks. Table 4 summarizes the task facet relationships with task effect results we found in analyzing text acquisition and state transitions on SERPs and content pages.

We looked at a number of other properties of basic eye fixation measurements and of characteristics of reading sequences identified by our reading algorithm. We also found that properties of reading subsequences within a page showed significant task effects. In particular, tasks affected the length of the reading sequence (in pixels) ($p = 0.001$), the average saccade to the next fixation in the reading sequence ($p < 0.05$) and the number of retrograde saccades in the reading sequence ($p < 0.05$). Retrograde saccades are regressions that are characteristic of reading sequences. There is evidence that the probability of a retrograde saccade is associated with the difficulty or unfamiliarity of the text (Starr and Rayner, 2001).
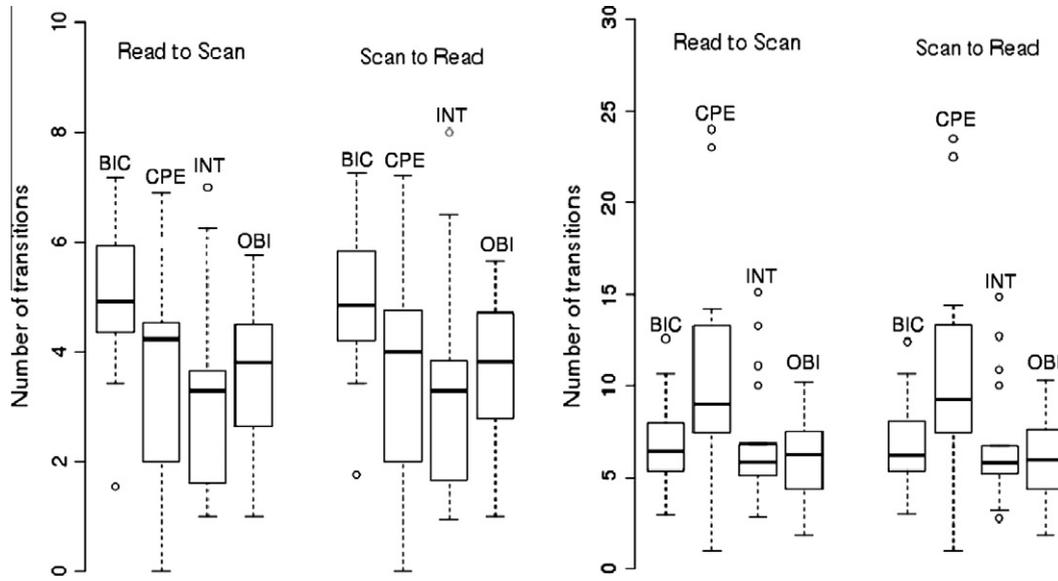
Fig. 4. Reading model state transitions per page. SERPs (left panel) and content pages (right panel).
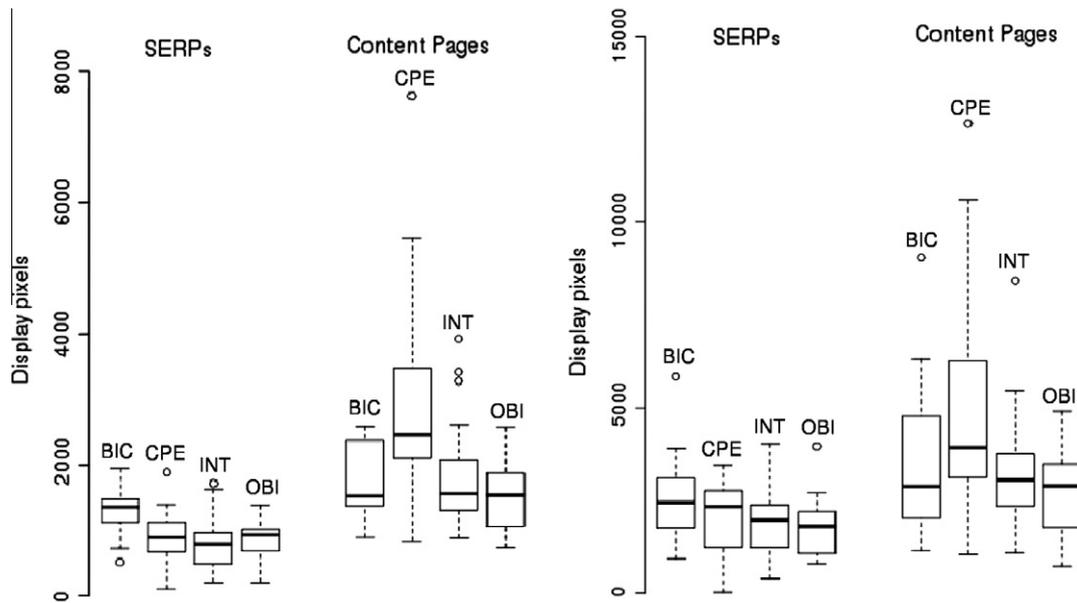

Fig. 5. Text acquired (pixel length) per page via scanning (left panel) and reading (right panel).

We also looked at the mean of the sum of the fixation durations on SERP pages and on content pages normalized by the number of pages. Diving this by the number of fixations in the task provides a 'cognitive speed' measure, which provides an indication of the relative duration of cognitive processing on each page type. SERPs had a 'cognitive speed' of about 1.5 times that of content pages. These results were consistent across tasks but there were no significant task differences. This result does agree with intuitions about the relative cognitive effort of processing a SERP page to recognize a potentially good link vs. processing a content page to extract information.

5.3. Individual biases in Reading models

Fig. 6 shows the differences between individuals for reading model state transition probabilities in the journalism and the genomics search tasks. For each participant, it shows the range of the reading model state transition probabilities that reflect a deci-

sion to switch from scanning to reading (Scan to Read) or to switch back to scanning rather than continue reading (Read to Scan). One can also see the differences in individual biases to make these decisions to switch. Note that the participants have been ordered by the mean transition probability measured over all of their tasks. One can see there are variations in the values of the Scan to Read parameter. When all participants are considered, it appears there is both a similar range of differences in values and the range of individual variability for the Scan to Read and Read to Scan transition probabilities for participants in both of the experiments. Individual differences in patterns of eye movements to acquire textual information have been demonstrated (Hyönä et al., 2002), so these results are not surprising.

What is interesting is the pattern for individuals in the state transitions. In Fig. 6, line segments have been drawn between the state transition values for each participant. These lines are coded to indicate the measured operation span for each partici-

Table 4
Summary of relationships between task effects and task facets in experiment 1 (journalism).

Variable	Related task facet(s)
Total text acquired on SERPs	Task complexity: more text acquired in BIC and OBI
Amount of text acquired (in reading states) per page on SERPs	Task goal: amorphous (BIC)
Text acquired and the number of state transitions per page on content pages	Task level: segment and task product: factual (CPE)

pant. For many participants, a higher mean Read to Scan transition probability is matched to a lower mean Scan to Read transition probability as compared to the rest of the participants. This is unexpected. To see why, consider the two state reading model in Fig. 1. The read and scan states are independent of one another, and it is obvious that the state transition probability for *Read to Scan* is not necessarily equal to the complement of the *Scan to Read* state transition probability. That is:

$$p \neq 1 - q \quad \text{and} \quad q \neq 1 - p$$

There is no analytic reason to expect this pattern in the data. As Fig. 1 shows, a user makes a choice at the end of a reading sequence of eye fixations, for example at the end of a line of text. They can continue reading by starting a new sequence of reading fixations, for example by reading the next line of text or jumping to a different place in a document but immediately begin reading. The alternative is to decide to scan (both reading and scanning, as defined here, are lexical and in each case the user can acquire information from the text they fixated on). In short, there is no reason to suppose a person's reading model biases could not include both a tendency to switch from scanning to reading and a bias to switch to reading from scanning. Fig. 6 provides evidence that for these very different task domains many participants were relatively consistent in their general bias towards reading or towards scanning. An appealing explanation is that an individual may have a general bias towards reading behavior and so their typical reading model parameters would differ from someone with a general bias towards scanning. Those who were inclined to switch to reading

from scanning were also inclined to continue reading as compared to others. In contrast, other participants were biased towards scanning in the same way.

Fig. 6 provides evidence for clear individual differences in reading models over two search task domains. There is a high correlation between the individual differences in each of the experiments. The non-parametric Spearman rho is 0.91 ($p < 0.001$) for the Journalism experiment and 0.85 ($p < 0.001$) for the genomics experiment. The results demonstrate there are differences in mean transition values and differences in individual variability in reading behavior. We observed a similar distribution of mean values for the Scan to Read and Read to Scan transition probabilities for individuals. These results suggest there may be a consistent individual bias towards reading or scanning regardless of the search topic domain.

For the journalism experiment results, we looked for correlations between these individual differences in state transition probabilities and the results of the operation span cognitive test administered before the experiment. No significant correlations were discovered. In a similar vein, the domain knowledge calculations for the individuals in the genomics experiment were compared with the individual differences shown in Fig. 6 (right hand panel), Again, no significant correlations were found.

6. Discussion

Reading is an important mechanism for information acquisition in interactive information tasks. To explore for relationships between factors known to affect information search behaviors and information acquisition during a task via reading, we implemented a model of reading based on eye movement patterns. We used the algorithm to process eye tracking data and describe the experiment task sessions in terms of the participant reading behaviors.

In the journalism experiment, we developed realistic work tasks to elicit generalizable task behaviors. Although we did so with an eye towards gaining some variation in the values of task facets in Li's categorization of task types, we did not seek to maximize the task facet differences in the various dimensions to avoid having tasks that were artificial. So the classifications of the experiment tasks have task facet differences, but also many similarities.
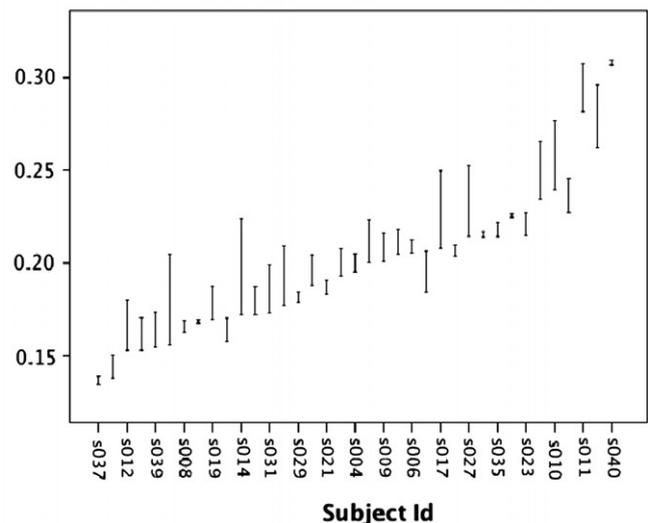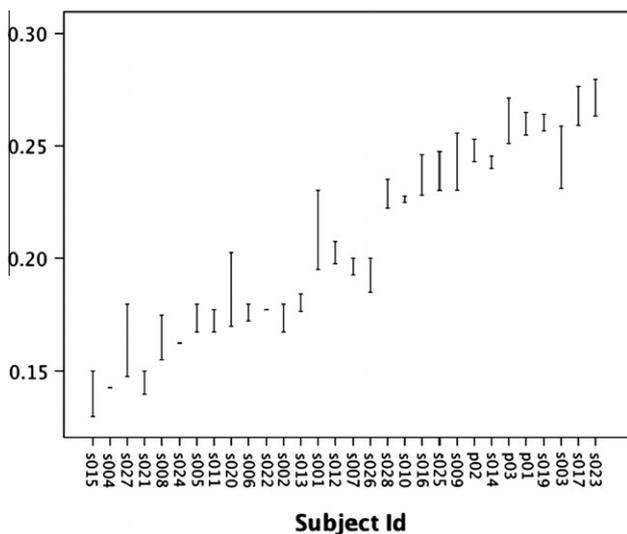

Fig. 6. Journalism (left) and genomics (right): differences between individual state transition probabilities *Scan to Read* (p) minus 1 − *Read to Scan* (1 − q) (mean over all tasks) (transition probability names (p, q) come from Fig. 1).

In view of this, it is interesting that effects seen at several levels of analysis of the low-level reading behaviors in patterns of eye movements seem to correlate well with some of the facet distinctions in the tasks (Table 1). Several of the effects are summarized in Table 4. For example, in the total amount of text acquisition to complete a task, the tasks with the complexity facet value of high (BIC and OBI), meaning more steps were required to complete the task, were distinguished from CPE and INT which had low complexity values.

The task influences on differences in reading behaviors were also manifest when the analysis was extended to page-level interactions in the SERP and content page types. For example, BIC, which was uniquely distinguished by having an amorphous task goal, appeared to have induced more intensive processing of SERPs as compared to the other tasks. In content pages, BIC and OBI tasks which had a task complexity facet value of high resulted in the highest total text acquisition. This can be attributed, as indicated by the complexity task facet value, to the number of steps required to meet the task need. When the activity on an average page was analyzed, the low complexity CPE task stood out as the task with the most activity, both in terms of text acquired and in the number of state transitions. CPE was distinguished from the other tasks in that it had a specific task goal, and a 'segment' task level, meaning that the task goal could be satisfied by processing only portions of a document. In particular, on average, CPE content pages had more state transitions and more reading, both in sequences and in isolated lexical fixations.

This task effect may, however, be due both to the nature of the task and to the relation of the task to the information system. It also serves to illustrate the complexity of the information interactions in this task environment and the value of the reading model methodology to illuminate information interactions at these fine-grained levels of analysis. CPE, being as it is specific and factual, is an example of a well specified search task. These types of search tasks generally allow people to structure more effective queries to the search system and receive SERPs more likely to hold highly-relevant results as compared to other task types. Appendix A has the task specifications as they were given to the participants. Contrast the type of query one might make for CPE and expected quality of search results with the same for BIC, where the nature of the information required in the task goal facet was amorphous. It seems plausible that participants processing pages for CPE may have visited those pages with the belief the needed task information was contained in those content pages. This observation about the expected search results in the SERP also helps to explain the greater activity on SERPs for the BIC task. It is also worth noting that our selection of highly attended pages for this analysis may have tended to select content pages reflecting some beliefs by the participants about the potential usefulness of the page to their task.

Nonetheless, the findings for the CPE task influence provide a good example of the value of our reading model methodology to investigate task effects on information acquisition interactions at multiple levels. We were able to investigate the total amount of text acquired, and look at interactions at the page level through the examination of average values of reading state switching and text acquisition for a page. It is obvious this approach can be extended to process sequences of pages, which constitute task segments, and so allow for research into information acquisition by task stage and in sub-tasks. This may be valuable in connecting identified correlations between eye movement patterns and task behaviors in user studies with real-world data because identification of entire tasks is often difficult in search logs and server-side data. It may be easier to identify sub-task behaviors.

A strength of this work is the diversity of the experiment domains and the participants. The information search tasks were conducted by distinct user groups: undergraduate journalism ma-

jors and a variety of undergraduates and graduate student in biological sciences. The types of tasks were developed in a situated work task setting. The undergraduate journalism majors were engaged in tasks that are typical for professionals (and for which they had already received training). The journalism study was designed to examine task behavior using realistic tasks in an unconstrained setting but in controlled conditions. The genomics study examined search task behavior using real search tasks for scientists. The students used a search system and accessed an appropriate part of the Medline database used by biological scientists. All of the tasks were examples of cognitively complex information search tasks a professional might conduct. The diversity of the participants and the nature of the tasks suggest our results can be generalized.

One unexpected result for both experiments was the peculiar distribution of participants ranked by their Scan to Read and Read to Scan state transition probabilities. An individual with a relatively higher Read to Scan state transition probability tends to also have, compared to the other participants, a lower Scan to Read state transition probability. There is no analytic explanation for this effect as the two states are independent of one another (Fig. 1). One rather intuitive explanation for this observation is that eye movement behaviors during text search interactions are primarily determined by individual differences and not much affected by the task type or nature of the content page, e.g. whether it is a SERP or a content page reached via click through on a SERP. This result provides support for the findings of Rayner et al. (2009) that there are individual and cultural differences in reading behaviors. Their work demonstrated that reading behaviors are notably different from other eye movement task behaviors, for example visual search for objects and face recognition. These results also support the findings of Hyönä et al. (2002) concerning differences amongst information processing eye movement patterns on texts.

The individual differences in reading behaviors we observed are a distinguishing feature of interactions with information systems involving text and may be a component in the complexity of interactions with these types of environments. This interesting pattern turns out to be rather robust. It holds at the experiment domain level where one might expect such a pattern to be unlikely given the experiment task treatments. At the task level, there is a clear pattern in the journalism tasks (Fig. 2), but not in the genomics domain tasks as might be expected since these tasks were of the same task type.

7. Conclusions

The availability of eye movement data in user studies of interactive information search provides a high-resolution view of the user's attention during these complex interaction tasks. Reading behavior is of particular interest because it is necessary for acquiring information from texts. It is clear reading is important for interactive information search tasks.

Reading models can capture aspects of an information acquisition process that is important for interactive information retrieval. Recall that our interest is in personalization of interactive information search. This means we need to provide input to information search processes such that the response from the system, say in the information resources offered, takes account of salient aspects of the user and their task. Understanding how reading models are affected by factors known to influence information search behaviors, such as the nature of the task and individual differences, provides a means to make richer models of tasks and users. The specific mechanism to discover the appropriate model parameterization is to look at correlations between the reading model parameters, in the present case state transition values, and

expressions of interaction behavior by users performing information seeking tasks.

State transition models of the reading/scanning eye movements reveal important information about a user's cognitive processing of displayed information. At the task level, users apply search and processing strategies that depend upon properties of their primary search task. In this work, we provide evidence for individual differences in reading models and for changes in an individual's visual cognitive processing strategy due to a high-level search task differences and web page effects.

Reading is a key visual cognition strategy used to meet the needs of high-level information needs in complex cognitive tasks. While there is substantial work showing a link between immediate cognitive tasks and visual cognition strategies (Hayhoe et al., 2007; Triesch et al., 2003) and individual differences in reading strategies (Hyönä et al., 2002), we are not aware of previous work showing effects of high-level information tasks on visual cognition strategies using reading models.

It is known eye movements are cognitively controlled (Findlay and Gilchrist, 2003). Our general interest is in the higher level information acquisition during information search that may be revealed through eye movements. In this sense, our work tries to bridge from modeling reading as an immediate cognitive process to modeling reading as a strategic cognitive process by asking how lower-level models of reading are influenced by factors that are known to influence higher-level cognitive processes that play out in complex information tasks over extended interaction sessions. It seems plausible that the user's task might be a strong influence in the decision to read or scan for information. Another possibility is that it is influenced by a mental state of the user, in particular the user's knowledge that is germane to the task. Yet another possibility is that individual differences, for example cognitive abilities, are the primary influence.

The approach we have taken in this paper is to look at eye movements in service of reading, that is, acquiring textual information. A contribution of this work is the development of reading models as a methodology to investigate information acquisition interactions in complex environments, specifically in interactive information retrieval setting. A virtue of the technique is the ability to look at several levels of interaction behaviors: over the entire task, in subsequences of tasks that may correspond to task stage, and in detailed analysis of interactions on types of pages and on individual pages. Since the algorithm requires only sequences of eye fixation coordinates and durations as inputs, our implementation of the EZ Reader model can be used to re-analyze eye tracking logs from existing experiments. It can also be used to build models of one aspect of the information acquisition process for users and tasks, which may be useful in personalizing information retrieval systems. The technique can also be used in practical applications. The algorithm is not computationally-demanding. Given eye tracking data, it could be employed in operational settings to make predictions about aspects of the user and their task by matching calculated reading models with constructed models of relationships between reading model-based information acquisition parameters and user and task behaviors.

In this work, we constructed reading models of an individual's information seeking session and fashioned a state model of the sequences of lexical fixations. Our focus was on two state transitions – *Scan to Read* and *Read to Scan*. Leaving aside transitions within the reading sequence (in particular the transition to the terminal fixation in the reading sequence) is justified by the statistical significance of the state changes. These two states also have intuitive appeal. It is natural to suppose at some level there is a decision – a commitment – to reading a text segment or just a single word.

Modeling reading in search task sessions, and more generally in complex information environments, involves the question: Which

cognitive level does the model address? This two state model is an expression of a hypothesis about the relation between reading and the use of reading as an information acquisition strategy. This is in contrast with lower-level reading models, such as EZ Reader, that describe the process of reading without getting into higher-level cognitive processes served by reading processes. A two-state model is a simple system and, at this level of cognitive analysis, about the simplest proposal that might possibly be fruitful. Certainly, this idea can be extended with multi-state models based on patterns of reading sequences and other dimensions, including page properties, cognitive load, and so on.

A general connection between information seeking tasks, at both high and low levels, and eye movement behavior seems plausible whenever information retrieval interactions depend on a visual interface. User attention and cognitive resources, such as working memory, constrain information acquisition for each interaction as well as the information carried across interactions in the session. So a user's information seeking strategy, say a search strategy, affects the visual cognition strategy, e.g. an initial stance to scan rather than read a search results page, because of user expectations about the information environment they will process.

A strong argument can be made for this expectation of task effects from previous research results. Information science research has established the existence of task effects on information search behaviors. Reading is essential to acquire information in text search environments and so it seems natural to expect task influences on the use of reading during information behaviors. It is gratifying, then, to see task influences on the total amount of text acquired, the number of state transitions on average pages and the average amount of text acquired on a page. Differential effects of tasks on reading in distinct page types, SERPs and content pages were all observed,

The fact that the observed task effects reflected differences in task facet values is also encouraging. It provides additional evidence for the usefulness of Li's task classification system. These results also support the general idea that interaction behaviors expressed in eye movement patterns and in other ways, for example by page dwell time, may be valid implicit indicators of the user's task type.

Acknowledgment

Appendix A. Tasks

A.1. Journalism study tasks

A.1.1. 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 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 2 years in the New York Times on this topic, and also in five other important newspapers, either US or foreign.

A.1.2. 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.

A.1.3. 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.

A.1.4. Copy editing (CPE)

Your assignment: You are a copy editor at a newspaper and you have only 20 min 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$4000 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 capita income to US$3000 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.

A.2. Genomics study tasks

The numbers refer to the TREC topic numbers in the 2005 genomics track. The categories are the MeSH tree names (e.g. G05), and the reference following the task title is the node in the MeSH tree. Task 22 was used for training. The other five tasks were used in the study.

A.2.1. Category I: genetic processes

7 DNA repair and oxidative stress (1.7. DNA Repair [G05.195])
Need: Find correlation between DNA repair pathways and oxidative stress.

Context: Researcher is interested in how oxidative stress effects DNA repair.

22 Relative response of p53 family members to agents causing single-stranded vs. double-stranded DNA breaks (1.4.2.1. Chromosome Breakage [G05.180.210.170])
Need: Does p53 respond differently to different DNA-damaging agents? Do they respond differently to single-strand vs. double-strand breaks?

Context: DNA damage may cause cell cycle arrest or apoptosis. p53 plays a role in mediating these sequelae of DNA damage.

A.2.2. Category II. genetic phenomena

45 Mental Health Wellness-1 (2.11. Linkage (Genetics) [G13.540])
Need: What genetic loci, such as Mental Health Wellness 1 (MWH1) are implicated in mental health?

Context: Want to identify genes involved in mental disorders.

42 Genes altered by chromosome translocations (2.18.4.3.12. Translocation, Genetic [G13.920.590.175.870]

Need: What genes show altered behavior due to chromosomal rearrangements?

Context: Information is required on the disruption of functions from genomic DNA rearrangements.

A.2.3. Category III: genetic structure

49 Glyphosate tolerance gene sequence (3.7. Genome [G14.340])
Need: Find reports and glyphosate tolerance gene sequences in the literature.

Context: A DNA sequence isolated in the laboratory is often sequenced only partially, until enough sequence is generated to identify the gene. In these situations, the rest of the sequence is inferred from matching clones in the public domain. When there is difficulty in the laboratory manipulating the DNA segment using sequence-dependent methods, the laboratory isolate must be re-examined.

2 Generating transgenic mice (3.7.1.5.37. Transgenes [G14.340.024. 340.825])
Need: Find protocols for generating transgenic mice.

Context: Determine protocols to generate transgenic mice having a single copy of the gene of interest at a specific location.

References

Belkin, N.J., 2008. Some (what) grand challenges for information retrieval. SIGIR Forum 42 (1), 47–54.

Bierig, R., Gwizdka, J., Cole, M.J., 2009. A user-centered experiment and logging framework for interactive information retrieval. In: Belkin, N.J., Bierig, R., Buscher, G., van Elst, L., Gwizdka, J., Jose, J., Teevan, J. (Eds.), Proceedings of the SIGIR 2009 Workshop on Understanding the User: Logging and interpreting user interactions in information search and retrieval. CEUR, Boston, MA. pp. 8–11.

Blanchard, H., Pollatsek, A., Rayner, K., 1989. The acquisition of parafoveal word information in reading. Perception and Psychophysics 46 (1), 85–94.

Borlund, P., 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. Information Research 8 (3).

Buscher, G., Dengel, A., van Elst, L., 2008a. Eye movements as implicit relevance feedback. In: Proceedings of CHI '08. ACM, Florence, Italy, pp. 2991–2998.

Buscher, G., Dengel, A., van Elst, L., 2008b. Query expansion using gaze-based feedback on the subdocument level. In: Proceedings of SIGIR '08. ACM, Singapore, pp. 387–394.

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

Conati, C., Merton, C., Muldner, K., Ternes, D., 2005. Exploring eye tracking to increase bandwidth in user modeling. In: Ardissomo, L., Brna, P., Mitrovic, A. (Eds.), Proceedings of User Modeling 2005. Springer-Verlag, Edinburgh, Scotland, pp. 357–366.

Drieghe, D., Desmet, T., Brysbaert, M., 2007. How important are linguistic factors in word skipping during reading? British Journal of Psychology 98, 157–171.

Drieghe, D., Pollatsek, A., Juhasz, B., Rayner, K., 2010. Parafoveal processing during reading is reduced across a morphological boundary. Cognition 116 (1), 136–142.

Duggan, G., Payne, S., 2008. Knowledge in the head and on the web: Using topic expertise to aid search. In: Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems. ACM, Florence, Italy, pp. 39–48.

Farzan, R., Brusilovsky, P., 2009. Social navigation support for information seeking. In: Houben, G.J. et al. (Eds.), UMAP 2009. Springer, Trento, Italy, pp. 66–77.

Findlay, J., Gilchrist, I., 2003. Active Vision: The Psychology of Looking and Seeing. Oxford University Press, New York.

Ford, N., Miller, D., Moss, N., 2005. Web search strategies and human individual differences: cognitive and demographic factors, Internet attitudes, and approaches. Journal of the American Society for Information Science and Technology 56 (7).

Freund, 2008. Exploring task-document relations in support of information retrieval in the workplace. Unpublished Dissertation. University of Toronto.

Granka, L., Joachims, T., Gay, G., 2004. Eye-tracking analysis of user behavior in WWW search. In: Proceedings of SIGIR '04. ACM Press, Sheffield, UK, pp. 478–479.

Guan, Z., Cutrell, E., 2007. An eye tracking study of the effect of target rank on web search. In: Proceedings of CHI '07. ACM, San Jose, CA, p. 420.

Gwizdka, J., 2008. Revisiting search task difficulty: behavioral and individual difference measures. Proceedings of the American Society for Information Science and Technology 45 (1), 1–12.

Hayhoe, M.M., Droll, J., Mennie, N., 2007. Learning where to look. In: Van Gompel, R.P.G., Fischer, M.H., Murray, W.S., Hill, R.L. (Eds.), Eye Movements: A Window on Mind and Brain. Elsevier, Oxford, pp. 641–660 (Chapter 30).

Hembrooke, H., Granka, L.A., Gay, G.K., Liddy, E., 2005. The effects of expertise and feedback on search term selection and subsequent learning. Journal of

the American Society for Information Science and Technology 56 (8), 861–871.

Hersh, W.R., Cohen, A.M., Yang, J., Bhupatiraju, R.T., Roberts, P.M., Hearst, M.A., 2005. TREC 2005 genomics track overview. In: Proceedings of the 2005 Text Retrieval Conference, pp. 14–25.

Hornof, A.J., Halverson, T., 2002. Cleaning up systematic error in eye-tracking data by using required fixation locations. Behavior Research Methods, Instruments, & Computers 34 (4), 592–604.

Hsieh-Yee, I., 1993. Effects of search experience and subject knowledge on online search behavior: measuring the search tactics of novice and experienced searchers. Journal of the American Society for Information Science and Technology 44 (3), 161–174.

Hyönä, J., Lorch, R., Kaakinen, J., 2002. Individual differences in reading to summarize expository text: evidence from eye fixation patterns. Journal of Educational Psychology 94 (1), 44–55.

Inhoff, A.W., Liu, W., 1998. The perceptual span and oculomotor activity during the reading of Chinese sentences. Journal of Experimental Psychology. Human Perception and Performance 24 (1), 20–34.

Inhoff, A.W., Rayner, K., 1986. Parafoveal word processing during eye fixations in reading: effects of word frequency. Perception & Psychophysics 40 (6), 431.

Juhasz, B., White, S., Liversedge, S., Rayner, K., 2008. Eye movements and the use of parafoveal word length information in reading. Journal of Experimental Psychology. Human Perception and Performance 34 (6), 1560.

Juvina, I., van Oostendorp, H., 2006. Individual differences and behavioral metrics involved in modeling web navigation. Universal Access in the Information Society 4 (3), 258–269.

Juvina, I., van Oostendorp, H., Karbor, P., Pauw, B., 2005. Towards modeling contextual information in web navigation. In: 27th Annual Meeting of the Cognitive Science Society, CogSci2005. Cognitive Society, Austin, TX, pp. 1078–1083.

Kambe, G., 2004. Parafoveal processing of prefixed words during eye fixations in reading: evidence against morphological influences on parafoveal preprocessing. Perception & Psychophysics 66 (2), 279–292.

Kammerer, Y., Gerjets, P., 2010. How the interface design influences users' spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. ACM, pp. 299–306.

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

Kelly, D., Belkin, N.J., 2004. Display time as implicit feedback: understanding task effects. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, Sheffield, UK, pp. 377–384.

Kelly, D., Cool, C., 2002. The effects of topic familiarity on information search behavior. In: Proceedings of the Second ACM/IEEE Joint Conference on Digital Libraries (JCDL '02). ACM, Portland, OR, pp. 74–75.

Kuhlthau, C.C., 1991. Inside the search process: information seeking from the user's perspective. Journal of the American Society for Information Science and Technology 42, 361–371.

Lawless, K., Shrader, P., Mayall, H., 2007. Acquisition of online information: knowledge, navigation and learning outcomes. Journal of Literacy Research 39 (3), 289–306.

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.

Li, Y., Belkin, N.J., 2008. A faceted approach to conceptualizing tasks in information seeking. Information Processing & Management 44 (6), 1822–1837.

Lin, S.-J., 2001. Modeling and supporting multiple information seeking episodes over the web. Unpublished Dissertation. Rutgers University.

Liu, J., Belkin, N.J., 2010. Personalizing information retrieval for multi-session tasks: the roles of task stage and task type. In: Proceedings of SIGIR '10. ACM, Geneva, Switzerland.

Liu, Y.-H., Wacholder, N., 2008. Do human-developed index terms help users? An experimental study of MeSH terms in biomedical searching. Proceedings of the American Society for Information Science and Technology 45 (1), 1–16.

Liu, W., Inhoff, A.W., Ye, Y., Wu, C., 2002. Use of parafoveally visible characters during the reading of Chinese sentences. Journal of Experimental Psychology. Human Perception and Performance 28 (5), 1213–1227.

Liu, C., Gwizdka, J., Liu, J., 2010. Helping identify when users find useful documents: examination of query reformulation intervals. In: Proceeding of the Third Symposium on Information Interaction in Context. ACM, New Brunswick, NJ, pp. 215–224.

Lorigo, L., Haridasan, M., Brynjarsdottir, H., Xia, L., Joachims, T., Gay, G.K., Granka, L.A., Pellacini, F., Pan, B., 2008. Eye tracking and online search: lessons learned and challenges ahead. Journal of the American Society for Information Science and Technology 59 (7), 1041–1052.

Morris, R., Rayner, K., Pollatsek, A., 1990. Eye movement guidance in reading: the role of parafoveal letter and space information. Journal of Experimental Psychology. Human Perception and Performance 16 (2), 268.

Oliveira, F.T., Aula, A., Russell, D.M., 2009. Discriminating the relevance of web search results with measures of pupil size. In: Proceedings of CHI '09. ACM, Boston, MA, pp. 2209–2212.

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), 801–823.

Pollatsek, A., Rayner, K., Balota, D.A., 1986. Inferences about eye movement control from the perceptual span in reading. Perception & Psychophysics 40 (2), 123–130.

Pollatsek, A., Rayner, K., Fischer, M., Reichle, E.D., 1999. Attention and eye movements in reading. In: Everatt, J. (Ed.), Reading and Dyslexia: Visual and Attentional Processes. Routledge, London, pp. 179–209.

Pollatsek, A., Reichle, E.D., Rayner, K., 2003. Modeling eye movements in reading: extensions of the EZ Reader model. In: Hyönä, J., Radach, R., Deubel, H. (Eds.), The Mind's Eye: Cognitive and Applied Aspects of Oculomotor Research, pp. 361–390.

Pollatsek, A., Reichle, E.D., Rayner, K., 2006. Tests of the EZ Reader model: exploring the interface between cognition and eye-movement control. Cognitive Psychology 52 (1), 1–56.

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

Rayner, K., Duffy, S., 1986. Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition 14 (3), 191–201.

Rayner, K., Fischer, M., 1996. Mindless reading revisited: eye movements during reading and scanning are different. Perception & Psychophysics 58 (5), 734–747.

Rayner, K., Pollatsek, A., 1989. The Psychology of Reading. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Rayner, K., Reichle, E.D., Pollatsek, A., 1998. Eye movement control in reading: an overview and model. In: Eye Guidance in Reading and Scene Perception. Elsevier, Oxford, pp. 243–268.

Rayner, K., White, S., Kambe, G., Miller, B., Liversedge, S., 2003. On the processing of meaning from parafoveal vision during eye fixations in reading. In: Hyönä, J., Radach, R., Deubel, H. (Eds.), The Mind's Eye: Cognitive and Applied Aspects of Oculomotor Research. Elsevier, Oxford, England, pp. 213–234.

Rayner, K., Li, X., Williams, C.C., Cave, K.R., Well, A., 2007. Eye movements during information processing tasks: individual differences and cultural effects. Vision Research 50 (21), 2714–2726.

Rayner, K., Smith, T.J., Malcolm, G.L., Henderson, J.M., 2009. Eye movements and visual encoding during scene perception. Psychological Science 20 (1), 6–10.

Reichle, E.D., Pollatsek, A., Fisher, D., Rayner, K., 1998. Toward a model of eye movement control in reading. Psychological Review 105, 125–157.

Reichle, E.D., Rayner, K., Pollatsek, A., 2004. The EZ Reader model of eye-movement control in reading: comparisons to other models. Behavioral and Brain Sciences 26 (04), 445–476.

Reichle, E.D., 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.

Reingold, E., Rayner, K., 2006. Examining the word identification stages hypothesized by the EZ Reader model. Psychological Science 17 (9), 742–746.

Roberts, P.M., Cohen, A.M., Hersh, W.R., 2009. Tasks, topics and relevance judging for the TREC genomics track: five years of experience evaluating biomedical text information retrieval. Information Retrieval 12 (1), 81–97.

Shannon, C.E., 1951. Prediction and entropy of printed English. Bell System Technical Journal 30 (1), 50–64.

Sherman, C., 2005. A new f-word for Google search results. Search Engine Watch.

Starr, M., Rayner, K., 2001. Eye movements during reading: some current controversies. Trends in Cognitive Sciences 5 (4), 156–163.

Taylor, R.S., 1968. Question negotiation and information seeking in libraries. College & Research Libraries 29, 178–194.

Taylor, A.R., Cool, C., Belkin, N.J., Amadio, W.J., 2007. Relationships between categories of relevance criteria and stage in task completion. Information Processing and Management 43, 1071–1084.

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 '08. ACM, London, pp. 152–159.

Toms, E., Mackenzie, T., Jordan, C., O'Brien, H., Freund, L.S., Toze, S., Dawe, E., MacNutt, A., 2007. How task affects information search. In: Workshop Pre-Proceedings in Initiative for the Evaluation of XML Retrieval (INEX), pp. 337–341.

Torralba, A., Oliva, A., Castelhano, M., 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 (4), 766–786.

Triesch, J., Ballard, D.H., Hayhoe, M.M., Sullivan, B.T., 2003. What you see is what you need. Journal of Vision 3, 86–94.

Vakkari, P., Hakala, N., 2000. Changes in relevance criteria and problem stages in task performance. Journal of Documentation 56 (5), 540–562.

Vakkari, P., Pennanen, M., Serola, S., 2003. Changes of search terms and tactics while writing a research proposal: a longitudinal research. Information Processing Management 39, 445–463.

Wen, L., Ruthven, I., Borlund, P., 2006. The effects on topic familiarity on online search behaviour and use of relevance criteria. In: Advances in Information Retrieval, 28th European Conference on IR Research, ECIR 2006. Springer, London, pp. 456–459.

White, R.W., Kelly, D., 2006. A study on the effects of personalization and task information on implicit feedback performance. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM'09. ACM, Arlington, VA, pp. 297–306.

White, S., Liversedge, S., 2005. The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. Psychonomic Bulletin & Review 12, 466–471.

White, R.W., Ruthven, I., Jose, J.M., 2005. A study of factors affecting the utility of implicit relevance feedback. In: Proceedings of the 28th Annual International Conference on Research and Development in Information Retrieval, SIGIR 2005. ACM, Salvador, Brazil, pp. 35–42.

White, R.W., Dumais, S.T., Teevan, J., 2009. Characterizing the influence of domain expertise on web search behavior. In: Proceedings of WSDM'09. ACM, New York, pp. 132–141.

Wildemuth, B.M., 2004. The effects of domain knowledge on search tactic formulation. Journal of the American Society for Information Science and Technology 55 (3), 246–258.

Zhang, X., Anghelescu, H., Yuan, X., 2005. Domain knowledge, search behavior, and search effectiveness of engineering and science students: an exploratory study. Information Research 10 (2), 217.