9-1-2013

Inferring User Knowledge Level from Eye Movement Patterns

Michael J. Cole
Rutgers University

Jacek Gwizdka
Rutgers University

Chang Liu
Rutgers University

Nicholas J. Belkin
Rutgers University

Xiangmin Zhang
Wayne State University, xiangmin.zhang2@wayne.edu

Inferring User Knowledge Level from Eye Movement Patterns

MICHAEL J. COLE[a], JACEK GWIZDKA[a], CHANG LIU[a], NICHOLAS J. BELKIN[a], XIANGMIN ZHANG[b]

[a] School of Communication and Information, Rutgers University, New Brunswick, NJ 08901
[b] School of Library and Information Science, Wayne State University, Detroit, MI 48202

Corresponding author MICHAEL J. COLE m.cole@rutgers.edu; JACEK GWIZDKA ipm2012@gwizdka.com; CHANG LIU changl@eden.rutgers.edu; NICHOLAS J. BELKIN belkin@rutgers.edu; XIANGMIN ZHANG xiangminz@gmail.com

Abstract The acquisition of information and the search interaction process is influenced strongly by a person's use of their knowledge of the domain and the task. In this paper we show that a user's level of domain knowledge can be inferred from their interactive search behaviors without considering the content of queries or documents. A technique is presented to model a user's information acquisition process during search using only measurements of eye movement patterns. In a user study (n=40) of search in the domain of genomics, a representation of the participant's domain knowledge was constructed using self-ratings of knowledge of genomics-related terms (n=409). Cognitive effort features associated with reading eye movement patterns were calculated for each reading instance during the search tasks. The results show correlations between the cognitive effort due to reading and an individual's level of domain knowledge. We construct exploratory regression models that suggest it is possible to build models that can make predictions of the user's level of knowledge based on real-time measurements of eye movement patterns during a task session.

Keywords knowledge detection, cognitive effort, user study, personalization, information search behavior

1. INTRODUCTION

While users frequently seek information about something they do not already know (Belkin, 2000), an essential aspect of all information behavior is that one must use existing knowledge to make progress towards the task goal. Identifying relevant knowledge and using it to guide search in a problem space is a fundamental aspect of models of cognition (Anderson, 1990 and Newell, 1990). Acquisition of new information is usually essential to the process of achieving the goal and it is typical for that new information to be acquired by reading.

The only way to acquire information visually is by repeated gaze on a location. One must also allocate attention during this process and there is still much that is unknown about how the process and mechanism of attention works (Wright & Ward, 2008). In reading, it has been shown that one can distinguish between eye movements that are engaged in information acquisition and those that are not (Rayner and Fischer, 1996 and Reichle et al., 2010). Research into reading eye movement patterns allows one to couple observations of eye fixations on words with the process of acquiring the meanings of the words (Rayner 1998). It is possible to not only understand what information the person has engaged with during search interactions, but also to learn important details of how it has been processed. Previous work shows this can be used to infer high level properties of the user's search situation, such as their current task type (Cole et al., 2010 and Cole et al., 2011b) and their experience of the difficulty of the task (Cole, Gwizdka, Liu & Belkin, 2011a).

Analysis of eye measurements is particularly attractive for study and modeling of information search behavior because research shows they are connected directly with mental states. In particular, the acquisition of the meaning of a word or phrase is revealed by real time measurements of eye fixations (Rayner, 1998 and Staub et al., 2010). Analysis of eye movement patterns, unlike other observations of information search behavior, allows for unmediated measurement of user mental states that have an essential role in the search process.

Our research goals include understanding the value of predicting domain knowledge to improve information retrieval system performance. For broad practical application, one needs to learn not only how to detect domain knowledge but also how that can be accomplished in a domain independent manner. Implicit detection techniques are desirable and so we have studied the problem of inferring domain knowledge by concentrating on observable behaviors.

Cognitive modeling of a person's existing knowledge is challenging. Consistent with the research showing that the time to acquire word meaning depends on existing concept knowledge (c.f Kieras, 1981 and Foss, 1982), the cognitive effort due to reading experienced by a user during search may be expected to reflect, in part, their existing domain knowledge. This paper presents results of a user study showing correlations between individual cognitive effort features and a person's level of domain knowledge. It also constructs regression models based on the cognitive effort features to explore the possibility of predicting a user's level of domain knowledge based on real-time measurements of their eye movement patterns.

2. RELATED WORK

2. 1 Implicit detection of domain knowledge during search

It is reasonable to think behaviors can be used to infer user knowledge levels without the need to consider the content of documents or queries. Domain knowledge or expertise has been shown to affect search behaviors. Users with low domain knowledge are found to have more non-productive queries (Allen, 1991), use less efficient concepts and make more query reformulation errors (Wildemuth, 2004), and resort to less effective search strategies (Hembrooke, Granka, Gay, & Liddy, 2005). Users with intermediate levels of knowledge about their task fail to select available documents of which they have the most knowledge as compared to users with high or low levels of domain knowledge (Cole, Zhang, Belkin, Liu & Gwizdka, 2011c). High level information search behaviors, such as dwell time, selected document ranks, and query length, have been used to construct a domain knowledge model (Zhang, Cole, & Belkin, 2011).

2.2. Eye movements in search

In text-based interactive information retrieval (IIR), information acquisition is mediated by eye movement patterns in service of the reading process. Eye movements are known to be cognitively-controlled (Findlay & Gilchrist, 2003). They provide a low-level behavioral observation of interactive tasks (Karn and Hayhoe, 2000 and Triesch et al., 2003) and are well-suited to represent the textual information acquisition process during search tasks.

Eye tracking has received considerable attention as a new source of data for research into the information search process. Much of the work in information science using eye tracking data has concentrated on eye fixations, for example to indicate which items are considered in ranked search results pages (Pan et al., 2007 and Brumby and Howes, 2008), or in identifying words useful for relevance feedback (Buscher et al., 2008a, Buscher et al., 2008b and Loboda et al., 2011). Eye tracking has identified patterns of processing documents, for example an "F" shape reading pattern for a search engine result page (SERP) (Granka, Feusner, & Lorigo, 2006). Granka et al., 2004 and Lorigo et al., 2008 studied the number of fixations, their duration and time on task in a user study of searches with the Yahoo! and Google search engines. Eye movements during different types of information retrieval activities have been investigated by looking at SERP interactions during informational and navigational tasks (Terai et al., 2008), different task types (Liu et al., 2010), information use (Cutrell & Guan, 2007), the effects of search page ranks on subsequent actions (Guan & Cutrell, 2007), and the usefulness of social navigation clues to users performing web searches (Loboda et al., 2011).

2.3. Eye movements and reading

Eye movement patterns are cognitively controlled and reading patterns have long been studied (Rayner, 1998). There are many results relating eye movements to semantic and cognitive processing states. Models of the reading process have been developed that explain observed fixation duration and word skipping behaviors.

The E-Z Reader model is a cognitively-controlled, serial-attention model of reading eye movements (Reichle, Rayner, & Pollatsek, 2004). It takes word identification, visual processing, attention, and control of the oculomotor system as joint determinants of eye movement in the reading process. The saccade (i.e., very fast movement of eyes during which eyes do not acquire any visual information) to the next word is programmed while the text in the current fixation is being cognitively processed.

There are several stages of text processing during fixations. In the E-Z Reader model it is supposed that the controller of eye-movement is triggered by completion of an early word identification stage, called the familiarity check, and the shift in attention to the next word selected takes place only after full lexical access is achieved. The mean minimum time to acquire the full meaning of a word is 151 ms (Reingold & Rayner, 2006) and the mean minimum time for the familiarity check is a little more than 110 ms (Pollatsek, Reichle, & Rayner, 2006) There is a labile period for reprogramming the pending saccade. If that time window is exceeded, the pending saccade will be executed when the cognitive processing of the current fixation is completed. Eyes remain fixated during the lexical processing period independently of the stimuli, for example even if the word is removed (Findlay & Gilchrist, 2003). The fixation duration depends on the familiarity and conceptual complexity of the text processed (e.g. Rayner & Duffy, 1986). The next saccade takes place only after cognitive processing is completed. This explains why observations of eye movements are connected directly with a person's mental states of information acquisition during search.

If the next word is recognized during the labile stage, the programmed saccade is canceled and a saccade to the next word is programmed. Frequently, a person is able to infer enough meaning to permit planning a saccade word target that is several words away from the current fixation. The E-Z Reader model does not account for higher-order cognitive processes and does not address language comprehension and conceptual processing, such as deductive or analogical reasoning (Reichle et al., 2004).

2.4. Reading eye movement pattern analysis and search

Buscher, Dengel, and van Elst (2008b) presents an eye tracking model of information acquisition processing by labeling sequences of fixations as reading that is more intense or engaged, 'reading', vs. a less engaged interaction, 'skimming'. Their reading model algorithm uses the position of the succeeding fixations to label the reading segments. They then show the labeled sequences can be exploited to select words that improve query expansion quality.

In previous work we have demonstrated relationships between eye movement patterns and task and page types (Cole et al., 2011b and Cole et al., 2010). We have implemented a model of reading based on empirical research into the reading process (Reichle et al., 2004) and used it to develop several measures of cognitive effort due to textual information acquisition. These cognitive effort measures are correlated with user task difficulty assessments and with objective task effort measured using high level behaviors, such as number of documents examined and document use (Cole et al., 2011a).

2.5. Eye movements and cognitive effort during reading

Reading process research has identified several observable indicators of cognitive effort associated with reading eye movements. Cognitive effort in reading arises from accessing word meanings, phrase and sentence parsing, and text comprehension. Each of these depend upon an individual's domain knowledge. Research has connected these aspects of processing effort with properties of reading eye movements.

Indications of reading effort can be inferred from eye movement patterns. An obvious one is reading speed. Reading speed will be greater if the text is easy to read (Rayner & Pollatsek, 1989), and is affected by word familiarity (Williams & Morris, 2004), words used in less frequently encountered senses (Sereno, O'Donnell, & Rayner, 2006), and when additional reflection is required to comprehend the concepts involved (Morris, 1994). Sentence parsing can also impact reading speed.

The cognitive processing needed to acquire word meaning and its meaning in context is also indicated by fixation duration. Rayner, Chace, Slattery, and Ashby (2006) show text comprehension processing has some observable effects in eye movements. Conceptually difficult text passages involved more fixations and slightly longer mean fixation duration.

As explained above, word skipping during reading sequences depends on the ability to recognize words. This controls fixation spacing. Fixation spacing is also associated with cognitive processing constraints. Perceptual span is the amount of text processed as a unit. Studies of reading in different orthographic systems show perceptual span describes a property of human cognitive processing. Fixation spacing while reading Chinese is about three characters (Tsai and McConkie, 1995 and Inhoff and Liu, 1998) compared with about 15 characters in languages like English. There are differences in perceptual span between phoneme-encoded systems, such as Hebrew, and character-based systems, e.g. English and Dutch. Pollatsek, Rayner, and Balota (1986) found that bilingual Hebrew and Dutch speakers changed their perceptual span when switching between the two languages. What is striking in these results is that across the orthographic systems, approximately the same number of concepts can be expressed in the different perceptual spans observed.

Carpenter and McDonald (2007) propose a model of neural-decision making for saccade programming that explains perceptual span as a function of competing cognitive mechanisms. Increased semantic processing requirements result in longer fixations. Increasing word familiarity increases the rate at which the mechanism reaches the next saccade decision threshold, but fixation proximity to the previous fixation increases the probability of executing the programmed saccade to the next word. The result is that when the fixation spacing is short, processing of even familiar words will tend not to increase the average fixation spacing. So a decreasing perceptual span is expected to correlate with unfamiliar words and conceptually difficult passages.

Reading sequences commonly include retrograde saccades, where the next eye fixation returns to a previous point in the text passage. It is a common feature of reading eye movement sequences and can have an incidence of 10–15% of the total fixations in the reading sequence (Boland, 2004). The number of regressions in a reading sequence, and the fixation durations of the regression fixation have been associated with conceptual complexity, the difficulty of reading passages, the resolution of words with ambiguous sense, and the reading goal (Rayner and Pollatsek, 1989 and Rayner et al., 2006). Domain experts are also observed to regress more frequently in text passages as compared to non-experts (Boland, 2004).

2.6. Summary

There is ample evidence that the user's level of domain knowledge affects their search behavior, so it is reasonable to think that observation of search behaviors can reveal the knowledge level of users. Eye movements are a low level behavior and their observable properties are determined by the user's interest and cognitive processing of the words. Of particular importance is the fact that eyes fixate until the meaning of the word(s) is acquired.

The cognitive nature of information interaction connects cognitive effort measurements derived from eye movements with user domain knowledge. The measures are closely associated with semantic processing, such as acquisition of word meaning, which in turn are correlated with a user's level of knowledge. Vocabulary knowledge is related directly to concept knowledge via internal representation of words indicating concept features or as the mechanism for accessing these concepts. This essential link between the meaningfulness of words and concept formation and use is a core aspect of research into the nature of concepts (c.f. Katz, 1972, Fodor, 1975, Armstrong et al., 1983 and Landauer, 2002). Despite the on-going difficulty of fixing the precise relationship between psycholinguistics and concept access and use, knowledge of vocabulary is well-accepted as an indicator of concept knowledge.

Observable eye fixations can be associated with semantic and cognitive processing during reading in ways that relate directly to cognitive effort. Word familiarity, sense disambiguation in context, and the conceptual difficulty of a text passage can all be related to eye movement features, specifically the duration and spacing of eye fixations, and fixation sequence patterns. Our previous work investigated relationships between the user's task type and transitions in reading strategies from scanning to extended reading and differences in the influence of page types (search results pages vs. content documents) on text acquisition and page processing when different tasks are being executed. This paper extends previous work by investigating eye movement-based measures of the cognitive effort due to reading, and showing correlations with the user's level of domain knowledge.

3. METHODOLOGY

3.1. Experiment and participant knowledge representation

3.1.1. ***Procedure and calculation of knowledge levels*** We conducted a user study to explore the effects of differences in domain knowledge on search behaviors. Undergraduate and graduate students (n=40) in biology-related programs were recruited to an on-campus usability laboratory. They read and signed a consent form and filled out a background questionnaire that solicited demographic information, computer experience and experience in search.

Domain knowledge representation is difficult. Asking people to self-assess their knowledge of some domain is difficult because of vague boundaries, understanding of the terms and projection of the scale for the assessment. How much difference is needed to say one knows a bit about a concept or has moderate knowledge? We approached this problem by selecting a comprehensive domain concept representation system and constructing a self-rating scale with 'bright line' anchors.

Medical Subject Headings (MeSH, http://www.nlm.nih.gov/mesh/) is a controlled vocabulary developed by the National Library of Medicine for indexing medical and biomedical literature. It contains over 25,000 concepts arranged in a collection of tree-like structures (tress with some entanglement). An attraction of the MeSH system is the detailed and comprehensive coverage over the broad biomedical domain, including biology, medicine, genetics, etc.

We used terms from three MeSH categories that corresponded to the search tasks used in the study. This use of rating controlled vocabulary terms has been employed previously in other studies to measure a user's domain knowledge (Zhang, Anghelescu, & Yuan, 2005) and proved to be an effective approach for the purpose of studying users of IR systems.

One strength of the study design is that participants made a series of independent judgments about their knowledge of specific concepts rather than a judgment overall about their knowledge of the general area (say genomics) or the task itself. We also asked participants to make those assessments as well, but do not use them in this paper. In other work, we compared the collection of specific MeSH concept judgments with their general assessment of domain and task knowledge and find correlations, but also significant variances (?).

In constructing the self rating scale we were concerned to have anchors the could elicit the same judgments from people and reduce the variance inherent in subjective assessments. One such anchor is clear – people can be relied upon to determine whether they know nothing about a specific concept. Likewise, it is reasonable to think that people can say when they have some idea of the concept without claiming they have a significant level knowledge of which they are confident. Relative levels of knowledge are more difficult to pin down. We chose to use a performance-like measure as another anchor. It seems plausible people can project whether they are confident enough about their knowledge of a concept to explain that concept to someone who is not an expert. While self-deception objections can be raised, this projected performance measure helps people to distinguish between having high knowledge of concept and feeling they have real expertise. Using these bright line anchors we fashioned a five point Likert scale (1–'No knowledge', 2– 'Vague idea', 3–'Some knowledge', 4–'High knowledge', 5–'Can explain to others'). The middle point of 'some knowledge' is not anchored although it can be confidently selected by exclusion of 'vague idea' and 'high knowledge'.

Before the experiment, participants rated their knowledge of 409 genomics-related MeSH terms from the three MeSH categories that corresponded to the search tasks used in the study. It took from 20 to 40 minutes to complete the rating process. This methodology of rating controlled vocabulary terms was used in a previous studies to measure a user's domain knowledge (Zhang et al., 2005) where it proved to be an effective approach for the purpose of studying users of IR systems.

The participant term ratings were processed to make a single measurement of their domain knowledge. The participant domain knowledge (PDK) was calculated as:

$$\text{PDK} = \frac{\sum_{i=1}^{m}(k_i * t_i)}{5 * m}$$

where k_i is the term knowledge rating and i ranges over the terms. m is the total number of terms rated (max=409) and t_i is 1 if rated and 0 if not. The sum is normalized by a hypothetical expert who rated all terms as 'can explain to others'.

The PDK measurements were then fashioned into a Euclidean distance matrix and clustered with Ward's hierarchical agglomerative clustering technique, which uses a minimum variance method to find compact, spherical clusters. Three well-distinguished levels of knowledge were identified in the participant group: high domain knowledge (HDK) (n=6), intermediate domain knowledge (IDK) (n=24), and low domain knowledge (LDK) (n=8).

After rating their domain knowledge the participants were led through a training task to gain familiarity with the search system. The training task was a very easy task taken from the 2004 TREC Genomics track. After completing the training task, participants were asked to do 4 out of 5 tasks, with two tasks alternated during the study. Before each task, the participant filled out a questionnaire asking about their familiarity with the task using a seven point Likert scale (1–'not at all', ... , 7–'extremely'). They were then given 15 minutes to perform the task. After each task there was a post-task questionnaire asking them to assess their experience of the search, including task difficulty and learning, using seven point Likert scales. An exit questionnaire was administered after they performed all of the tasks. Each participant was paid $25 for completing the experiment.

3.1.2. ***Tasks*** The search tasks were taken from the 2004 TREC Genomics track (Hersh et al., 2005), which were ad hoc retrieval tasks from 50 topics relating to five general types (Roberts, Cohen, & Hersh, 2009). 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 & Wacholder, 2008).

A simulated work task approach (Borlund, 2003) was used to design the task presentation to the participants. The tasks were presented without changes from the TREC track. Participants were asked to find and save all of the documents useful for answering the task questions. Five tasks were used in the experiment. The tasks, with TREC topic numbers noted, as presented to the participants were:

Category I: Genetic Processes

7 DNA repair and oxidative stress

- **Need:** Find correlation between DNA repair pathways and oxidative stress.
- **Context:** Researcher is interested in how oxidative stress affects DNA repair.

Category II. Genetic Phenomena

45 Mental Health Wellness-1

- **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

- **Need:** What genes show altered behavior due to chromosomal rearrangements?
- **Context:** Information is required on the disruption of functions from genomic DNA rearrangements.

Category III: Genetic Structure

49 Glyphosate tolerance gene sequence

- **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

- **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.

Each participant did four tasks. All of the participants completed the tasks 2, 7, and 45. Twenty participants did topic 42 and the other twenty did task 49. The presentation order of the tasks was randomized in a blocked Latin Square design. We switched the tasks halfway through the experiment because we were interested in eliciting a range of information behaviors looking for knowledge effects and task 42 was observed to be too easy for the participants for this purpose. The tasks were presented in a blocked and counter-balanced design with task 49 substituted for task 42 at the half-way point of the study. In this paper we do not distinguish between tasks in the analysis, but the substitution of tasks is nonetheless a limitation of the experiment. The questions presented to the participants were the TREC genomics track descriptions including the need and context, as shown above.

We implemented a search system using Indri from the Lemur toolkit (www.lemurproject.org). 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 to 2004 period (n = 1.85 million) to allow for reasonable retrieval performance. Search interactions were recorded using a multi-source logging system (Bierig, Cole, Gwizdka, & Belkin, 2010). Fig. 1 shows screen shots of the search system presentation of search results and the content links, which were article abstracts. During the search, all of the participants' interactions with the computer system, including eye gaze, were logged using a Tobii T-60 eye tracker (1280 × 1024 @ 60 Hz). We used eye fixation data as calculated by the Tobii Studio software. Technical reasons prevented analysis of two participants.

3.2. Eye movement data processing

3.2.1. *Representing the user's experience of information acquisition due to reading* We implemented a line-oriented reading model based on the E-Z Reader model (Reichle et al., 2004) and used the algorithm to process the location and duration of participant eye fixations as

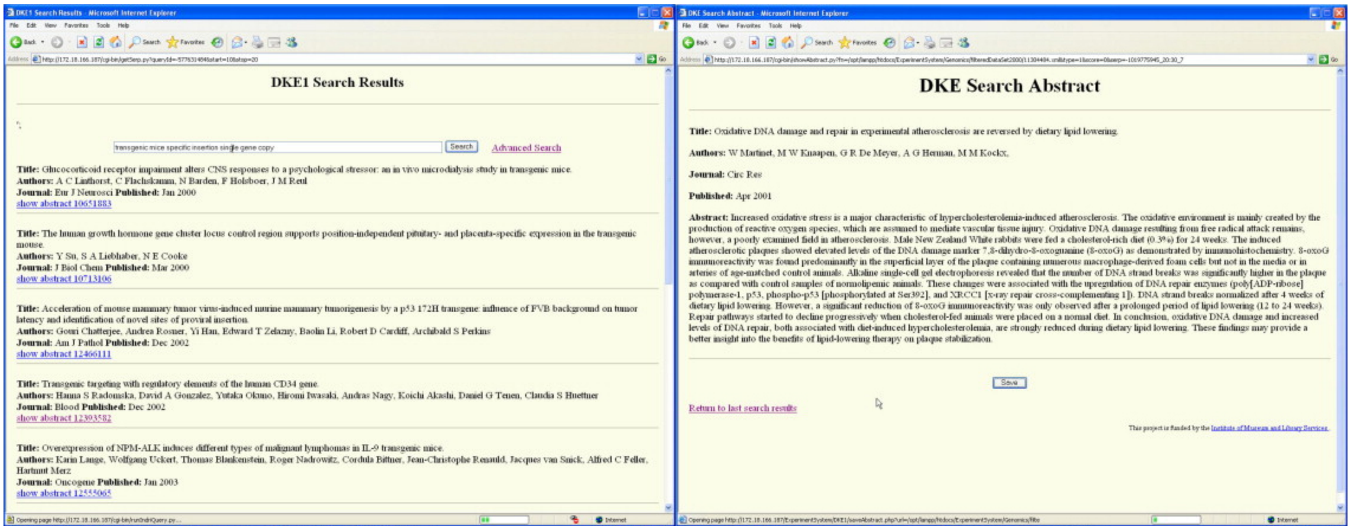

Figure 1 Examples of user study search results and content pages

logged by the Tobii eye tracker. First we used a fixation duration of 113 ms to classify which eye fixations were highly likely to result in word meaning acquisition. The reading model was then used to group these lexical fixations into reading sequences (figure 2). These reading sequences are taken to represent the user's experience of information acquisition via reading.

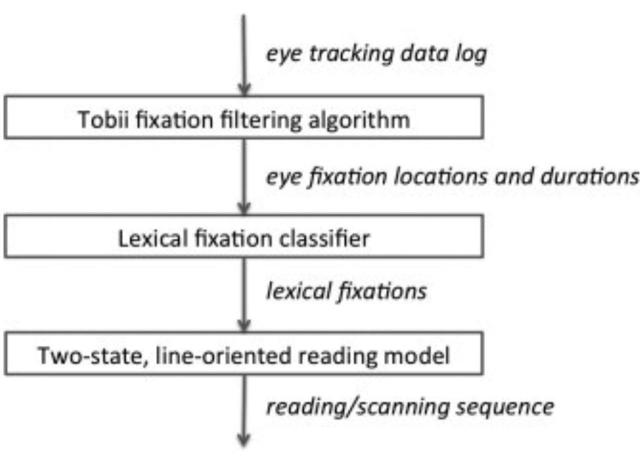

Figure 2 Eye movement data processing to extract reading sequences

Our algorithm for classifying reading sequences is line-oriented. Eye movement data from a user reading several lines of text in a sequence would generate a representation of several reading sequences. This is a limitation of our algorithm. additionally, our algorithm does not attempt to learn font sizes and physical line spacing on a page to decide if two reading sequences should be unified. Much of information search reading, however, consists of rather shorter units of text engagement, at the level of phrases and sentences.

3.2.2. *Calculating cognitive effort due to reading* Several cognitive effort measures based on reading fixation sequences have been developed (see section 2.5 for details). Studies of reading employ these measures but they have not been construed by others as cognitive effort, per se. The measures are:

- fixation duration in excess of the minimum required for lexical processing,
- the existence and number of regression fixations in the reading sequence,
- the spacing of fixations in the reading sequence, and
- reading speed, as defined as the length of text acquired per unit time.

Lexical access duration excess (LADE)

Fixation duration is an indicator of the cognitive processing required to establish the meaning of the word, and the meaning of the word in context.

A minimum of 113 ms is required to acquire the meaning of a word (Pollatsek et al., 2006). The lexical access duration excess (LADE) is the additional observed fixation duration beyond this lexical minimum.

Regressions

A regression fixation is a fixation that, in left to right reading, returns to a portion of the text already processed. We operationalize a regression measurement as a count of the regression fixations in a single reading sequence consisting of at least four fixations.

Perceptual span

Perceptual span reflects the spacing of the fixations in the horizontal dimension while reading and describes the amount of text one takes in as a unit. We measure the mean perceptual span in a reading sequence as a cognitive effort feature. Since regression fixations may occur, a reading sequence can have several left to right subsequences.

$$perceptualspan = \frac{\sum_j \left(\sum_i |f_{i+1}(x) - f_i(x)| \right) / n}{m}$$

where:

f is a fixation in a reading sequence,

$f_i(x)$ is the fixation x coordinate. For top down text orientation use the y coordinate.

m is the number of left to right reading subsequences in the reading sequence, and $j = 1,...,m$.

n is the number of fixations in the left to right reading subsequence and $i = 1,...,n - 1$.

The reading/skimming distinction made by Buscher et al. (2008b) is similar to our definition of mean perceptual span, although we classify reading sequences in a different way.

Reading speed

Reading speed is the ratio of the amount of text processed (the *reading length*) to the processing time. Reading speed is a function of the duration of the individual fixations in the reading sequence, the spacing of the fixations (perceptual span), and the regressions in the reading sequence.

3.3. Reading sequence level analysis of the user's experience of information acquisition

In this paper, we represent the user's experience of information acquisition due to reading during the search as a contiguous collection of reading sequences. Using the operationalization described above, cognitive effort measures were calculated for each reading sequence.

A reading sequence that consists of many fixations might have particular significance for analysis because of the extended attention a user has allocated to the text. As noted in section 2.5, fixation duration is associated with the cognitive processing needed to acquire word meaning and meaning in context. At the same time, multiple fixations on a line of text result in taking in not only the meaning of the words but also the concepts that are referenced using the words. Recall also that it is common for some fixations in an extended reading sequence to be retrograde. Such fixations are associated with conceptually difficult passages, resolution of references, and domain expertise.

In the analysis, general statistics on all reading sequences are examined. We then focus on a subset of longer reading sequences. It seems plausible these longer reading sequences may be more indicative of domain knowledge effects because of the extended attention to that text by the user. A reading sequence that consists of many fixations might mark points where the participant was more likely to be using concepts during information acquisition and one can hypothesize greater concept use is associated with increasing number of fixations in a reading sequence. It also indicates points in the task session where a person may more likely have the experience of acquiring information because of the attention allocated to acquiring that text.

For practical reasons, we used four or more fixations as a threshold to select the reading sequence subset because any eye fixation regressions are unlikely to be meaningful for sequences of less than four fixations. That is, for a sequence of three or fewer fixations it is difficult to say if a retrograde fixation was really a return to a previously processed word or just an isolated scanning fixation. This four fixation threshold is arbitrary, however, and it could have been set higher. We found that the distribution of reading sequences by length followed a power law, and significantly fewer sequences would be available for analysis if the number were set much higher. Future work will look at differences in the cognitive effort measurements as a function of reading sequence length to see if knowledge effects are present.

3.4. Modeling domain knowledge using eye movement behaviors

We want to make models of domain knowledge from the eye movement behavior data in order to explore for behavioral features and interactions between features that can contribute to eventual development of effective prediction models that can be used for personalization. For these first steps of exploratory modeling, each reading sequence in the study is treated as an observation of the participant's domain knowledge. A supervised learning approach is employed and the models were learned from the cognitive effort vectors for each reading sequence labeled with the PDK measurement (3.1.1) for the participant.

We constructed two exploratory models: a simple linear model and a model using the random forests ensemble technique. In both cases, we explore to see if the models can indicate the knowledge level of a given participant using the following procedure. First, just the reading sequences are selected, ignoring the isolated fixations, for each *task session* by participant. These *task session* reading sequences are then used as input to the model to calculate a predicted domain knowledge value for that participant. This procedure is carried out for each of the four tasks. Then the mean of the four domain knowledge values for each participant is calculated. This mean value is then compared to the participant's MeSH domain knowledge (PDK) to see if they are correlated.

Our goal is to do exploratory modeling to validate the model approaches and isolate significant behavior features, rather than make predictions. The evaluation procedure, in particular, is not appropriate for testing a true predictive model. We used all of the cognitive effort data to construct the models, however the knowledge levels are learned without consideration of the user or the task. Since these both varied, the comparison of the knowledge level calculated by the models given the reading sequences for a single participant task session (and then averaged over the four task sessions completed by the participant) with the PDK calculated independently from the MeSH concepts is a valid way to understand the relative classification efficacy of the models.

3.4.1. *Linear model construction* For each of the long reading sequences, a vector of the cognitive effort measures was constructed. Each vector was labeled with the PDK value. A Gaussian family linear model was then constructed using this bag of labeled cognitive effort vectors.

3.4.2. *Random forests* Random forests (RF) provide a more sophisticated modeling approach to our problem. RF is an ensemble machine learning technique that learns both the model structure and the parameters from the data, rather than proposing a specific model of the variables for fitting, as in the linear regression approach (Breiman, 2001). RF models are resistant to overfitting and provide performance comparable to gradient boosted decision trees (Mohan, Chen, & Weinberger, 2011), and SVMs (Breiman et al., 2003).

Model performance for RF is measured by the out of bag (OOB) error. OOB error is very close to the cross validation error because each tree in a random forest is constructed by random resampling of the data and also the variables, selecting the most discriminative feature at each node. At each step roughly 37% of the data is held out to test the tree (Breiman et al., 2003) in the process of finding the best variable to split on from the random subset of the variables selected for consideration. The rest of the data is used to calculate the OOB error.

3.4.3. ***Suitability to user-centered interaction models*** Random forests have become an important modeling technique for situations where there are fewer observations than predictors. In the present work, there are more observations than features. The attractive aspects of RF for our work are rooted in the complexities of human interactions with information during search. There are inherent limitations in handling interactions that are conditionalized by non-local factors, for example the overall goal of a task session, the immediate information goal, and the information seeking strategy employed for the current search stage. Purely local techniques, such as linear regression and similar models, may be particularly ill-suited to models of cognitively-driven information seeking.

Random forests is an appropriate exploratory modeling approach given our lack of understanding of the causal mechanism by which knowledge is used during reading. RFs do not require pre-selection of features and they have good predictive performance even if many features used are irrelevant to the classification. This is well-suited to our user-centered representation of the information acquisition process, which is both a model of information search and a cognitive modeling application. Another nice characteristic of fitted RF models is their ability to use all data but automatically adjust the influence of outliers because of the random sampling of features and data. Our exploratory modeling of cognitive effort effects during search cannot easily dismiss outliers because we cannot know if they may be significant observations of cognitive processing. They may be exceptions due to instrumentation or cognitive glitches during the experiment that are not indicative of the influence of knowledge on the reading process.

Information search in task sessions involves non-linear and conditionalized relationships (Wildemuth, 2004, Qiu, 1993 and Liu and Belkin, 2010). It is reasonable to suppose such relationships may affect the cognitive effort during the task session. A complex and evolving structure of feature importance may be a basic characteristic of high-fidelity task session models. Decision tree ensemble methods are attractive because they exploit subspace partitioning that reflect interaction effects. Individual trees in the forest can capture conditional structure expressed in the data because each tree classifies in a subspace and will correspond to a distinct feature set that may be correlated with the conditional structure of the object of the model (Bengio, Delalleau, & Simard, 2010) (See Bengio (2009) for a cogent review of learning deep cognitive architectures).

Random forests have other attractive features for modeling cognitive environments. They are insensitive to monotone transformations of subsets of features, again because of the random forests exploitation of random subspaces. In our case, the cognitive effort features, while measuring different cognitive constructs, may well have correlations between them because they all depend on fixation location and duration. For example, perceptual span is an average of saccade distances in the reading sequence and can be expected to be correlated with the ratio of reading length to reading sequence total duration. Such feature subcollections may covary in the same direction but have differences in covariance that depends on the conditionalized situation for the information acquisition process. This may be a general problem for model production and analysis in cognitive approaches to learning human information interactions.

Finally, random forests modeling can handle mixtures of continuous and categorical features. This is well-suited to combining multiple behavior observation sources of search interaction. In this work we use only continuous cognitive effort measures, but are extending it to incorporate more traditional behavioral features, including categorical measures such as whether a page has been revisited or was used to click though to another document.

From a number of perspectives, RF approaches are well-suited to model cognitive mechanisms that are unknown and, for now, unobservable. They are a good match for the general problem of modeling human search information interaction from a user-centered perspective.

Random forests is an appropriate exploratory modeling approach for domain knowledge given our lack of understanding of the causal mechanism by which knowledge manifests itself in information acquisition through the reading process. We used the RandomForest library in the statistical computing environment, R, which implements Breiman (2001).

4. RESULTS

4.1. Eye movement behaviors and domain knowledge

When all reading sequences are considered, the distribution of each of the cognitive effort measures was not normal, therefore non-parametric methods were used to analyze the data. Correlations were found between participant domain knowledge and three of the cognitive effort measures: Perceptual span (Kruskal-Wallis $\chi^2(36) = 4734.25$, $p < .001$), median LADE (Kruskal-Wallis $\chi^2(36) = 5570.10$, $p < .001$), and reading speed (Kruskal-Wallis χ^2 $(36) = 105.09$, $p < .001$).

Long reading sequences might better reflect concept use by participants during information acquisition because of the attention allocated to acquiring that text (see section 3.3). The idea is that these are points in the task session where the user is more engaged with the content and concepts of the text. It is therefore more likely knowledge effects may be detected in the cognitive effort measurements for those reading sequences. For practical reasons, we used four or more fixations as a threshold to select these reading sequences which comprised 7.5% (19477/258586) of the total collection of reading sequences. The mean number of regressions in this subset was reasonably close to a normal distribution but was not correlated significantly with domain knowledge.

The other cognitive effort measures for long reading sequences were correlated with domain knowledge. Perceptual span was normally distributed and correlated (ANOVA $F(1, 19199) = 29.14$, $p < .001$), as was reading speed (ANOVA $F(1, 150) = 5.23$, $p = .024$). Median LADE was not normally distributed but was still correlated with domain knowledge (Kruskal-Wallis χ^2 $(36) = 4724.89$, $p < .001$).

4.2. Linear model results

4.2.1. ***Linear model construction*** Using all of the long reading sequences, a linear model was constructed using all of the reading sequence cognitive effort vectors. The contribution and significance of each of the cognitive effort features is shown in table 1.

Table 1 Linear model of domain knowledge by cognitive effort

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.3859	0.0082	47.08	< 2e-16
numRegressions	0.0050	0.0012	4.26	2.1e-05
perceptualSpan	-0.0001	< 0.0001	-3.14	0.0017
readingLength	0.0001	< 0.0000	8.29	< 2e-16
readingSpeed	0.0877	0.0156	5.62	< 2e-08
maxDur	< 0.0001	< 0.0001	4.54	< 6e-06
medianDur	-0.0002	< 0.0001	-10.06	< 2e-16
totalDur	< -0.0001	< 0.0001	-0.19	0.8517

To see if this simple model could indicate the knowledge level of a given participant, we selected just the reading sequences, ignoring the isolated fixations, for each task session by participant. These reading sequences were used as the model input to calculate a domain knowledge value for that participant. This procedure was carried out for each of the four tasks. The mean of the four domain knowledge values was calculated for each participant and compared with the self-rated MeSH domain knowledge (PDK). The knowledge levels calculated by the model were correlated with the PDK MeSH values (ANOVA $F(1,36) = 4.78$, $p=0.035$). The standard deviation of the mean domain knowledge value produced by the model was an order of magnitude lower as compared to the standard deviation of the PDK value (0.011 vs. 0.132).

To examine the model performance, we grouped the participants by the mean domain knowledge value produced by the model using the hierarchical cluster technique describe in 3.1.1. These groupings were compared to the knowledge levels identified in the MeSH-term rating groups (table 2). The linear model gave reasonable discrimination between the high and low domain knowledge groups. After removing non-native English speakers (n=14), the model results were better for the high knowledge group, but deteriorated somewhat for the low knowledge group (Table 3).

Table 2 Classification errors: All participants

	Linear model prediction		
PDK groups	low	inter	high
low	2	4	2
intermediate	5	12	7
high	0	2	4

Table 3 Classification errors: Native English only

	Linear model prediction		
PDK groups	low	inter	high
low	0	2	1
intermediate	3	9	5
high	0	0	4

4.3. Random forests modeling results

Like the linear model, the RF model was trained on all of the reading sequences and then used to calculate a domain knowledge value for each participant using just the data from the four task sessions carried out by the participant.

The relative significance of each of the features in the RF model is shown in table 4. The relative importance of each cognitive effort feature (%IncMSE) is expressed as the amount of error reduced by splitting the tree on the feature. This is a kind of information gain measure. The number of regressions offer relatively little gain compared to some of the other cognitive effort measures

Table 4 Random forests model of domain knowledge: Cognitive effort feature importance

	%IncMSE	IncNodePurity
numRegressions	29.43	8.73
numRegressionsThreshold Length4	27.87	8.85
perceptualSpan	54.12	55.94
readingLength	69.10	56.23
readingSpeed	62.25	56.39
maxDur	41.18	37.95
medianDur	64.52	43.70
totalDur	62.93	52.65

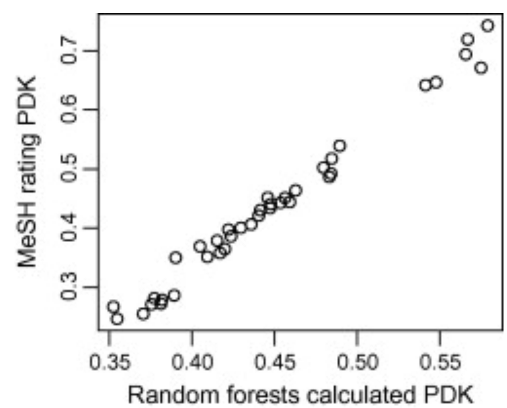

Figure 3 Random forest model domain knowledge correlation with MeSH knowledge based on long reading sequences

We test the RF model in the same way as for the linear model. A domain knowledge value was calculated using the model for each participant task session. Then the mean of these domain knowledge values for each participant's four task sessions was calculated. Figure 3 shows the high correlation between the participant's MeSH domain knowledge (PDK) and the RF knowledge calculation (ANOVA F (1,36) = 3913.3, $p < .001$). The calculated domain knowledge values were clustered in the same way the MeSH domain knowledge representation was clustered and the domain knowledge groups (high, intermediate, low) can be compared (Table 5). The classification accuracy is excellent and is perfect when only native English speakers are included (Table 6).

Table 5 Classification errors: All participants

	Random forests group		
PDK groups	low	inter	high
low	8	0	0
intermediate	1	23	0
high	0	0	6

Table 6 Classification errors: Native English only

	Random forests group		
PDK groups	low	inter	high
low	3	0	0
intermediate	0	17	0
high	0	0	4

4.4. Model differences by level of domain knowledge

Continuing in the exploratory spirit, it is interesting to inquire into the relative importance of the cognitive effort features in the random forests model. Figure 4 shows the relative importance of cognitive effort features in random forests models, measured by the % increase in mean square error, fitted to the three domain knowledge groups. One can see that the number of fixations and regressions in a reading sequence contribute relatively less to the fitted models as compared to mean perceptual span, reading speed and length, total lexical fixation duration, and LADE measures (maxDur, medianDur) on the constitutive fixations. The pattern of relative feature importance does not vary much over the fitted models.

5. DISCUSSION

It is intuitively plausible that user characteristics that influence the process of information search should be reflected in both high level behaviors, such as document dwell time, document use, and query formulation patterns, and in low level behaviors associated with the information acquisition process such as eye movement patterns. These user properties may be constraints or key parameters in the process of formulating and carrying out a strategy and tactics to acquire the information needed to complete the task.

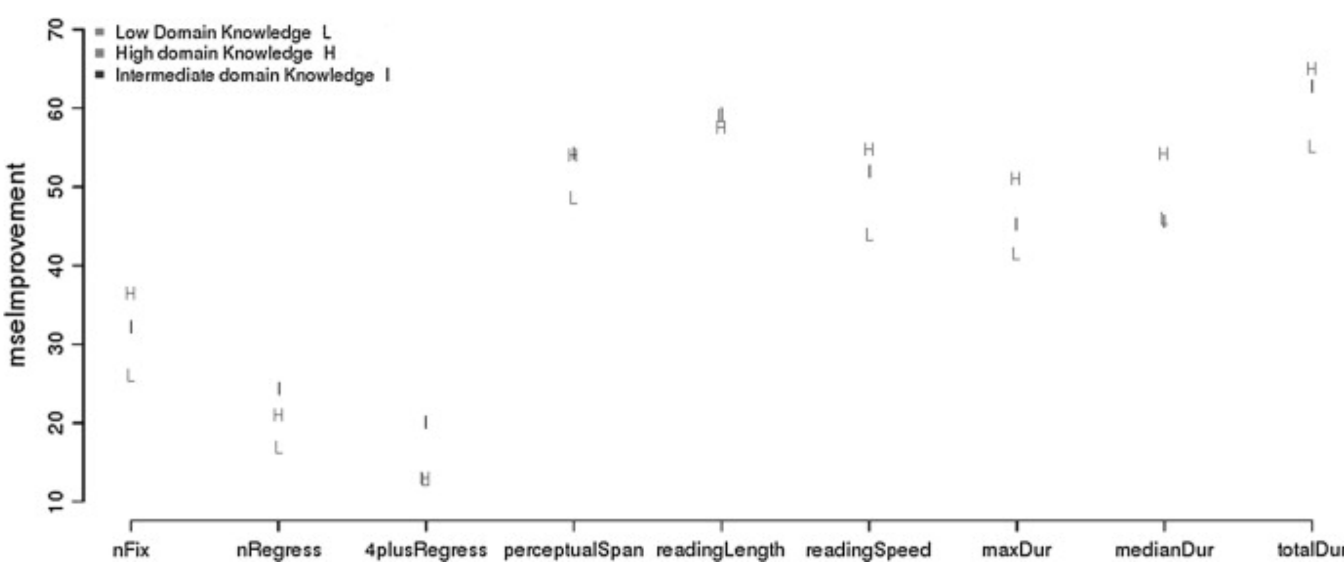

Figure 4 All participants: RF model feature importance

The low level eye movement behaviors we studied show correlations between user domain knowledge and measurements of eye movement patterns that reflect cognitive effort in the reading process. Reading speed, the fixation duration in excess of that minimally needed to acquire word meaning (LADE), and the spacing of fixations (perceptual span) were all significantly correlated with the user's genomics domain knowledge.

These results are consistent with the claim that the cognitive effort measures can represent the user's semantic experience during search which provide a basis for decision-making and other actions as exemplified by higher-order behaviors. The correlations with user-assessed domain knowledge and the success of the preliminary domain knowledge models supports this claim because we expect the user's knowledge to manifest itself in the process of both selecting text to read and in the acquisition of word meaning and concepts from the text. To be sure, the relationship between knowledge and this interaction process is complex. It is to be expected that experts will select text differently than non-experts in accordance with their understanding of the task goal. We believe a strength of grounding our approach in the measurement of objective cognitive effort avoids many complexities for analysis of search interaction in context because it is not necessary to posit a particular model of the unobservable cognitive process that results in the observable actions. Apart from the challenge of formulating such models in the absence of a general model of human cognition, any such model is likely not testable in complex interactions characteristic of real world search. Instead, we have chosen to focus on modeling the *process* of information acquisition because it is observable and because it has a one-to-one relationship with an aspect of the user *experience* of search interaction. The causal linkage between the user's knowledge, cognitive abilities, and their interaction with text in the context of their search goal, enables a new approach to investigate details of the search interaction process. Further, there is reason to think this approach is robust given our success in applying it in independent user studies to different constructs: task difficulty and domain knowledge.

The exploratory modeling was carried out with a desire to validate the model approaches and isolate significant behavior features, rather than make predictions. All of the data was used to make the models, although we learned the knowledge level without consideration of the user or the task.

The linear model did a reasonable job of classifying users against the MeSH term knowledge representation. Model prediction was improved when we took account of the native language of the participant. Such an improvement is expected since reading in a second language might be expected to affect the user's cognitive effort in reading processes somewhat independently of their level of domain knowledge. The improvement in the model classification performance when this source of noise in the relationship between level of domain knowledge and reading effort is eliminated is further evidence of the validity of the cognitive effort modeling approach.

Performance was much better when we used random forests, a more sophisticated modeling technique. Again, the level of domain knowledge was learned using each reading sequence as an observation of domain knowledge. However, the domain knowledge level was a unique characteristic of the user because no two users had the same domain knowledge value, and the performance of the random forests model may therefore be somewhat misleading as a predictive model. Future work will explore performance of the model in the prediction task.

Random forests models were constructed using cognitive effort observations from the three domain knowledge groups identified by hierarchical clustering. The results show little difference in the structure of the models by domain knowledge. There were fewer high domain knowledge (n=6) and low domain knowledge (n=8) participants compared to the intermediate domain knowledge (n=24) participants, and so the uncertainty around the feature importance measures is greater for those models. Still, the overall agreement of the relative feature importance is quite good. The stability of the model structure when fitted to different knowledge groups helps to explain the observed success of the random forests model in discriminating between the domain knowledge groups when the model is learned from all participants.

It is interesting to see that a number of the cognitive effort features had roughly equal contribution to the model (table 1). This suggests good predictive models are likely to have significant complexity.

The reading model analysis technique requires only input of recent eye fixations, specifically the location and duration of the fixation, and simple processing. These measures of the user's current processing of text meaning could be generated in near real-time and can be available for every interaction segment in the task session. Personalization of search interaction would benefit most from early prediction of the user's domain knowledge, so one direction for further analysis is to identify when a domain knowledge prediction can be made with reasonable confidence in a task session.

Our examination of the eye tracking logs shows that reading in the search sessions was unlike extended reading of a text. It was characterized by reading short snippets in one section, e.g. in search results, followed by very short reading sequences that might be characterized as scanning and isolated fixations for sufficient duration to acquire word meaning. Our methodology has some shortcomings in this regard. We do not, for example, model extended reading sequences of multiple lines of text. If a person read a paragraph that had four lines, they would be classified as a sequence of four reading sequences. Our algorithms can be improved in this regard, but differences in font sizes and layouts on different content pages introduce complexities that need to be accommodated. Similar challenges exist in calculating how much text has been read, and so on. It is important to note that application of the deep body of work on eye movements and reading to problems in dynamic information search environments must proceed cautiously because of the differences in the experiment environments and the main problems researchers seek to address. Our approach has been to concentrate on work related to processing at the sentence, phrase, and word level. We have been especially cautious in applying reading research work directed at more abstract levels, such as text comprehension, and work at structural levels of texts and presentation, such as how people interact in different page layouts precisely because translation to search environments is unlike reading a page in a book. How to apply existing research into reading in information search task sessions is an important problem and needs to be addressed experimentally.

5.1. Limitations

For both the high level and low level models, this work has been carried out in one experiment and domain. In future work we will apply the same techniques to an independent user study where participants carried out journalism tasks.

Domain knowledge representation is difficult to achieve and there are a number of shortcomings in the MeSH-based representation used in this study. For example, the coverage of rated MeSH terms (n=409) in the entire MeSH space (n=25,186) is an important limitation. This reflects a general problem for knowledge representation. It is exceedingly hard to say how much coverage in a concept space is needed for a useful representation. An attraction of the MeSH system is the detailed and comprehensive coverage in the domain. Unfortunately, such resources are not available for other domains. We believe one strength in our study design as used is that participants made a series of independent judgments about their knowledge of specific concepts rather than a judgment overall about their knowledge of the general area (say genomics) or the task itself. In other work we compared the collection of specific concept judgments with the general assessment (Cole et al., 2010) and found correlations, but also significant variances.

For the eye movement pattern cognitive effort model, the analysis in this paper concerns a sequence of reading eye fixations. We assume that all of the reading fixations represent engaged reading and are not a 'mindless' passing of the eyes over the words. Eye movement behaviors are likely to be naturally aligned with a user's experience of search as a sequence of page interactions. It is plausible domain knowledge effects on search results pages may be different than on the link (content) pages. Such page level analysis is an obvious next step since it correlates with the high level page-oriented observational units, such as document use and dwell time.

Our ultimate goal is to be able to build systems that can automatically personalize search interactions. To address the problem of general domain and user application of such a system we have focused on learning user and other models from observed behaviors of interaction with information during search. One thread in this paper is the suggestion that analyzing eye movement patterns to make moment to moment cognitive effort measurements can provide a direct and robust means to observe low level behaviors that connect directly with user cognitive states. An important question then is the general applicability of eye movement patterns analysis. Dyslexia has significant prevalence (~5%) in the population and may have a deep association with disruptions in eye movement patterns, perhaps in difficulties in phonological processing of seen text or in an impaired capacity to monitor eye movement (Lallier and Valdois, 2012 and Rayner et al., 1995). In our study we did not explicitly screen for these disorders and it is possible our data set included individuals with these conditions. In the larger project of making personalized systems it is clear significant attention will need to be given to accommodation of individuals with such disorders. There is already work on the effects of dyslexia on information search (MacFarlane et al., 2010). While this is a clear limitation in our study, it could be a significant attraction for further development of cognitive-effort eye movement-based personalization because detection of such users may be more difficult with less direct behavior observations.

6. CONCLUSIONS

The main contribution of this paper has been to demonstrate the plausibility of building models to infer a user's level of domain knowledge using certain information behaviors that require no need to process the content of queries or documents. Such models could potentially be domain-independent. Another contribution of this paper is the presentation of a technique based on analysis of eye movement patterns to represent the user's textual information acquisition process. From this one can model an important aspect of the user's *experience* of interactive information retrieval. We presented two exploratory regression models to infer user domain knowledge from eye tracking logs. Implicit detection of the user's domain knowledge or knowledge of their current task would be useful not only for search systems but also across a wide variety of information systems, for example intelligent tutoring systems.

Eye movement patterns are especially powerful for development of user-centered information systems because they have a direct relationship with the cognitive processing that connects document content with the user's knowledge of the meaning of the text. Patterns of spatial-temporal processing of regions of pages, or transitions from processing information objects to system interactions offer promising features for modeling the user's cognitive engagement in the search process. There is a wealth of research in cognitive psychology and related fields that provide a solid empirical foundation for analysis of eye movement patterns in information search interactions.

The technique we have developed is not necessarily restricted to research laboratory settings. The eye movement pattern methodology has very low computational demands. It involves trivial calculations using the position and duration of fixations over a few seconds. Calculations of cognitive effort due to reading can be made easily while the person is engaged in their task, enabling real-time detection of the user's level of domain knowledge.

While the results are from a single study and knowledge domain, it is plausible that these low-level information behaviors reflect cognitive effort and search strategies that depend on a user's knowledge of their task domain. Future work will make prediction models based on cognitive effort data from this user study and perform a similar analysis on results from an independent user study in the journalism task domain.

A person's use of knowledge during search drives the interaction process. These results provide evidence that the level of a user's domain task knowledge can be inferred from eye movement observations available during their search process. This could allow for dynamic personalization of an information system to improve its effectiveness.

7. ACKNOWLEDGEMENTS

IMLS grant LG-06-07-0105-07 supported this work. We also thank the anonymous HCIR 2011 workshop reviewers for comments on an early version as well as the anonymous reviewers of this paper for their useful suggestions.

8. REFERENCES

Allen B. (1991). Topic knowledge and online catalog search formulation. *The Library Quarterly* 61: 188–213

Anderson J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Armstrong S., Gleitman L., & Gleitman H. (1983). What some concepts might not be. *Cognition*, 13: 263–308

Belkin N.J. (2000). Helping people find what they do not know. *Communications of the ACM*, 43:58–61

Bengio Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1): 1–127

Bengio Y., Delalleau O., & Simard C. (2010). Decision trees do not generalize to new variations. *Computational Intelligence*, 26: 449–467

Bierig R., Cole M.J., Gwizdka J., & Belkin N.J. (2010). A data analysis and modelling framework for the evaluation of interactive information retrieval. In Gurrin C. (Ed.) et al., *Proceedings of ECIR 2010*, (Milton Keynes, UK, 2010), 5993: 673–674

Boland J.E. (2004). Linking eye movements to sentence comprehension in reading and listening. In Clifton C., Carreiras M. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*. Taylor & Francis: 51–76

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

Breiman L. (2001). Random forests. *Machine Learning*, 45: 5–32

Breiman L., & Cutler A. (2003). *Setting up, using, and understanding random forests V4.0*. University of California, Department of Statistics.

Brumby D.P., & Howes A. (2008). Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection. *Human–Computer Interaction,* 23: 1–46

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

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

Carpenter R., & McDonald S.A. (2007). LATER predicts saccade latency distributions in reading. *Experimental Brain Research*, 177: 176–183

Cole M.J., Gwizdka J., Bierig R., Belkin N.J., Liu J., Liu C., et al. (2010). Linking search tasks with low-level eye movement patterns. In *Proceedings of the 17th European Conference on Cognitive Ergonomic*s. Delf, The Netherlands

Cole M.J., Zhang X., Liu J., Liu C., Belkin N.J., Bierig R., et al. (2010). Are self-assessments reliable indicators of topic knowledge? *Proceedings of the American Society for Information Science and Technology conference*, ASIS&T, Pittsburg, PA

Cole M.J., Gwizdka J., Liu C., & Belkin N.J. (2011). Dynamic assessment of information acquisition effort during interactive search. *Proceedings of the American Society for Information Science and Technology conference*, ASIS&T, New Orleans, LA

Cole M.J., Gwizdka J., Liu C., Belkin N.J., Bierig R., & Zhang X. (2011). Task and user effects on reading patterns in information search. *Interacting with Computers*, 23: 346–362

Cole M.J., Zhang X., Belkin N.J., Liu C., & Gwizdka J. (2011). Knowledge effects on document selection in search results pages. *Proceedings of SIGIR 2011*, ACM, Beijing, China: 1219–1220

(2007). What are you looking for? An eye-tracking study of information usage in web search. *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM: 416

Findlay J., & Gilchrist I. (2003). *Active vision: The psychology of looking and seeing*. Oxford University Press

Fodor J.A. (1975). *The language of thought*. Cambridge, MA: Harvard Univ Press

Foss D. (1982). A discourse on semantic priming. *Cognitive Psychology*, 14: 590–607

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

Granka L.A., Feusner M., & Lorigo L. (2006). Eye monitoring in online search. In *Passive eye monitoring*. Springer: 347–372

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

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: 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 *The fourteenth text retrieval conference (TREC 2005)*. Gaithersburg, MD: NIST: 14–25

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: 20–34

Karn K.S., & Hayhoe M.M. (2000). Memory representations guide targeting eye movements in a natural task. *Visual Cognition*, 7: 673–703

Katz J.J. (1972). *Semantic theory*. New York: Harper & Row

Kieras D. (1981). Component processes in the comprehension of simple prose. *Journal of Verbal Learning and Verbal Behavior*, 20: 1–23

Lallier M., & Valdois S. (2012*). Sequential versus simultaneous processing deficits in developmental dyslexia*. InTech.

Landauer T.K., (2002). On the computational basis of learning and cognition: Arguments from LSA. In N. Ross (Ed.), *The psychology of learning and motivation*, 41: 43–84

Liu J., Cole M.J., Liu C., Belkin N.J., Zhang J., Gwizdka J., et al. Search behaviors in different task types. *Proceedings of JCDL 2010*, ACM, Gold Coast, Australia (2010)

Liu J., & Belkin N.J. (2010). Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. *Proceedings of SIGIR 2010*, 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: 1–16

Loboda T., Brusilovsky P., & Brunstein J. (2011). Inferring word relevance from eye-movements of readers. *Proceedings of the 15th International Conference on Intelligent User Interfaces*, ACM:. 175–184

Lorigo L., Haridasan M., Brynjarsdóttir H., Xia L., Joachims T., Gay G. et al. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59: 1041–1052

MacFarlane A., Al-Wabil A., Marshall C.R., Albrair A., Jones S.A., Zaphiris P. (2010). The effect of dyslexia on information retrieval: A pilot study. *Journal of Documentation*: 307–326

Mohan A., Chen Z., & Weinberger K.Q. (2011). Web-search ranking with initialize gradient boosted regression trees. *Journal of Machine Learning Research: Workshop and Conference Proceedings*, 14: 77–89

Morris R. (1994). Lexical and message-level sentence context effects on fixation times in reading. Journal of Experimental Psychology: *Learning, Memory, and Cognition*, 20: 92

Newell A. (1990). *Unified theories of cognition: The William James lectures 1987*. Cambridge MA: Harvard University Press

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: 801–823

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

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

Qiu L. (1993). Markov models of search state patterns in a hypertext information retrieval system. *JASIST*: 413–427

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

Rayner K., Pollatsek A., & Bilsky A. (1995). Can a temporal processing deficit account for dyslexia? *Psychonomic Bulletin & Review*: 501–507

Rayner K., Chace K.H., Slattery T.J., & Ashby J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Readin*g, 10: 241–255

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

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

Rayner K., & Pollatsek A. (1989). *The psychology of reading*. Mahwah, New Jersey: Lawrence Erlbaum

Reichle E., Reineberg A., Schooler J. (2010). Eye movements during mindless reading. *Psychological Science*, 21: 1300–3010

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: 4–22

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: 445–476

Reingold E., & Rayner K. (2006). Examining the word identification stages hypothesized by the EZ Reader model. *Psychological Science*, 17: 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 systems. *Information Retrieval*, 12: 81–97

Sereno S.C., O'Donnell P., & Rayner K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, 32: 335

Staub A., White S.J., Drieghe D., Hollway E.C., & Rayner K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 35: 1280–1293

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. *Proceedings of IIiX'08*, ACM, New York: 152–159

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

Tsai C., & McConkie G. (1995). The perceptual span in reading Chinese text: A moving window study. In *Seventh International Conference on the Cognitive Processing of Chinese and Other Asian Languages*, Hong Kong.

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

Williams R., & Morris R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *Journal of Cognitive Psychology*, 16: 312–339

Wright R., & Ward L. (2008). *Orienting of attention*. Oxford University Press

Zhang X., Cole M.J., & Belkin N.J. (2011). Predicting user domain knowledge from search behaviors. *Proceedings of SIGIR 2011*, ACM, Beijing

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