Natural Search User Interfaces for Complex Biomedical Search: An Eye Tracking Study

Ying-Hsang Liu , Paul Thomas, Marijana Bacic , Tom Gedeon & Xindi Li

Published online: 10 Aug 2017.

Submit your article to this journal ↗

Article views: 13

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Check for updates

Natural Search User Interfaces for Complex Biomedical Search: An Eye Tracking Study

Ying-Hsang Liu[a,b] [iD], Paul Thomas[b,c], Marijana Bacic[a,d] [iD], Tom Gedeon[b] and Xindi Li[e]

[a]School of Information Studies, Charles Sturt University, Wagga Wagga, Australia; [b]Research School of Computer Science, The Australian National University, Canberra, Australia; [c]Microsoft, Canberra, Australia; [d]Monash Health, Melbourne, Australia; [e]Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA

ABSTRACT

Controlled vocabularies such as Medical Subject Headings (MeSH) have been extensively used to organise information resources in the biomedical domain. However, the usefulness of these terms for information access has not been rigorously evaluated in interactive search environments. The objective of this study was to gain an understanding of domain experts' interactions with novel search interfaces within the context of biomedical information search, with a goal of better interface design of information retrieval systems. An eye tracking study of biomedical domain experts' interactions with novel search interfaces was conducted, considering user's individual differences and gaze behaviour. The findings suggest that types of search interfaces have significant effects on gaze behaviour in terms of fixation-based measures of areas of interest, i.e. visual attention to the elements of title, author, abstract and MeSH terms in document surrogates. Significant interaction effects between cognitive style and user interaction with search interfaces were found, specifically the amount of attention to MeSH terms by analytic and wholistic searchers. These findings contribute to our understanding of the relationship between cognitive styles and gaze patterns in information search. Based on these findings, the implications of individual differences and gaze behaviour for search interface design are discussed.

Introduction

Controlled vocabularies such as MeSH (Medical Subject Headings) have been extensively used to organise information resources in the biomedical domain. The usefulness of controlled vocabularies for information access, however, has not been extensively evaluated in interactive search environments (Behnert & Lewandowski, 2017; Golub et al., 2016; Shiri, 2012). Current search systems such as PubMed and MEDLINE, that are based on MeSH, use various techniques (e.g. query auto-completion and suggested term mapping) to map user queries to potentially useful query terms and relevant documents. This mapping to MeSH terms functionality is a feature of the search user interfaces produced by Wolters

Kluwer Health and EBSCO as evident in the Ovid and EBSCOhost platforms, respectively, in relation to the MEDLINE database that is offered by both publishers to libraries on a subscription basis.

From a practical perspective, some platforms have been set to enable the default search options on offer to map the queries to MeSH terms automatically. Successful searching of the MEDLINE database is more likely if the MeSH terms are used in the construction of search queries. For example, when searching MEDLINE on Ovid, the query of *bad breath* is mapped to the Subject Headings of Halitosis, Adult, Sulfur Compounds, Adolescent, Middle Aged, Oral Hygiene, Tongue, Smoking, Periodontal Diseases and Young Adult for users to select. In the PubMed database, typing the first three letters of the query *bad breath* produces two candidate terms *badminton* and *bad breath* through the auto-completion feature. When the query term *bad breath* is selected, the query is mapped to the MeSH term Halitosis automatically and translated into the following query syntax, which is a combination of searches using MeSH terms and free text searching across all search fields: 'halitosis'[MeSH Terms] OR 'halitosis'[All Fields] OR ('bad'[All Fields] AND 'breath'[All Fields]) OR 'bad breath'[All Fields]. In some hospital libraries, the library management has enabled the subject term mapping feature as default. These groups of users would have been exposed to the importance of MeSH and would possibly seek a high degree of relevance of the document being retrieved by search. A case in point is the Clinicians Health Channel portal (http://www.health.vic.gov.au/clinicians), accessible to all public hospital employees in Victoria, which provides access to MEDLINE (Ovid search interface) and other electronic resources. MEDLINE's default search is the Advanced Search with the map to subject terms feature enabled. As the resources provided by the Victorian Department of Health and Human Services are aimed to assist clinicians in locating evidence in order to provide quality care, and MEDLINE access maps to subject terms when accessed through the portal, this suggests clinicians would benefit from their searches directly mapping to MeSH terms. However, what has motivated the decision to make this the default setting is unknown and it is unclear whether this was the decision by the Department of Health and Human Services, Victoria, the vendor contracted to provide the service (Wolters Kluwer Health and Medical Director) or if any librarians were possibly consulted about it.

From a research perspective, however, the usefulness of these techniques for search user interface design in support of query reformulation has rarely been evaluated in interactive search environments (Hearst, 2009, 2011; Wilson, 2011). User studies have suggested that domain experts benefit the most from MeSH (Hersh, Haynes, & McKibbon, 1994; Liu & Wacholder, 2017) but that they also need search tools to support their formulation and reformulation of queries (Lykke, Price, & Delcambre, 2012; McKibbon & Fridsma, 2006). The design of a single search box in modern search systems presumes that users have very specific questions when they look for information. It is assumed that these questions can be translated into queries that match the terms in the documents with relevant information. However, empirical studies have demonstrated that users have problems formulating queries to represent their questions and reformulating queries in their interactions with search systems (Belkin, Marchetti, & Cool, 1993; Tamine & Chouquet, 2017; Wacholder, 2011). It is also recognised that user queries are typically very short representations of complex information needs (Jansen, Spink, & Saracevic, 2000). Since user interactions with search systems are construed as an inherently interactive process (Belkin et al., 1993), the

design of search interfaces for query reformulation tasks is important in support of search interactions.

The objective of this research was to gain an understanding of the domain experts' interactions with novel search interfaces within the context of biomedical information search, with a goal of better interface design of information retrieval (IR) systems. Specifically, we investigated how individual differences, such as levels of domain knowledge, search experience and cognitive styles affect search processes and eye gaze, with or without the MeSH terms as feedback.

Related Work

Query Formulation/Reformulation Support

Since the beginning of online intermediary searching in the 1970s, controlled vocabularies have been designed for query formulation/reformulation tasks. Searchers can look up controlled vocabularies assigned to documents in search indexes for query formulation tasks, and use these terms for reformulating queries during search processes. From a system design perspective, researchers have used the techniques of relevance feedback (Koenemann & Belkin, 1996), real-time interactive query expansion (White & Marchionini, 2007) and a hybrid approach that includes query logs (Agosti, Cisco, Di Nunzio, Masiero, & Melucci, 2010) to support the user's query formulation/reformulation tasks. At the level of search user interfaces, researchers have proposed various techniques to help users select potentially relevant documents and navigate the search results, such as the visualisation of document inter-relationships (Swan & Allan, 1998), explicit term distribution information (Hearst, 1995) and search results navigation (Leroy, Xu, Chung, Eggers, & Chen, 2007; Mu, Lu, & Ryu, 2014).

However, very few studies have explicitly considered the user's cognitive aspects of search interactions with suggested search terms for refining queries (cf. Liu & Belkin, 2008; Liu & Wacholder, 2017; Niu & Kelly, 2014). As indicated in a study of user modelling, the implicit feedback from clicks and query reformulations in Web search (Joachims et al., 2007), it is important to obtain detailed evidence of the user's query reformulation decision-making process from user search behaviours, such as search terms, time spent, clicks and eye tracking data, because user search behaviour data can be used to improve overall search experiences.

User Characteristics

Within the interactive IR research framework, some research has focused on the relationship between user characteristics and search performance (Ruthven & Kelly, 2011). The user characteristics of domain knowledge and search experience have attracted researchers' attention partly because they are assumed to be relevant for the training of expert searchers (e.g. Fenichel, 1981; Hsieh-Yee, 1993). Cognitive abilities and personality characteristics are associated with how people process information; cognitive styles – an individual's preference or tendency to process information – have also been extensively studied in the context of user interactions with IR systems (Ford, 2015; Heinstrom, 2010).

Domain knowledge refers to an individual's level of expertise in a subject discipline. Research in this area has focused on the relationship between the level of expertise and search performance, but there are mixed results (Hsieh-Yee, 1993; Pao et al., 1993). For instance, it was found that there is no relationship between clinical knowledge and search effectiveness in medical students' MEDLINE online searches (Pao et al., 1993). It was found that domain experts can perform significantly better than search experts when MeSH terms are available in the search interface (Liu & Wacholder, 2017). While ordinary searchers may be satisfied with the search results from short queries, effective search in specialised domains requires the domain knowledge and ability to formulate complex queries (Leroy et al., 2007). However, the search results from current IR systems have not been optimised for domain experts by considering their domain expertise and querying behaviours.

Search experience refers to a searcher's skills in interacting with IR systems. Liu's (2010) meta-analysis of eight user studies in controlled experimental settings suggested that search experience has a small and overall positive effect on search performance in terms of recall measures, with an effect size of .04. Studies of the relationship between search experience and search process variables revealed that searchers with online database experience use more thesaurus terms than inexperienced searchers, and free-text searching is associated with lack of online searching experience (Fenichel, 1981; Howard, 1982). Overall, these studies suggest that search experience with specific search systems may affect the use of system features, such as thesaurus terms.

Cognitive style is an individual's tendency or preference to process information in parts or in wholes (Peterson, Deary, & Austin, 2003). The influence of cognitive abilities and personality characteristics on a person's interaction with information can be summarised as '[T]he whole is more than the sum of the parts' (Heinstrom, 2010, p. 4). The wholistic-analytic dimension of cognitive style is of interest to IR researchers since it is considered an important factor affecting user search behaviour and search performance. For instance, research findings have suggested that people with different cognitive styles prefer different kinds of interface design for Web directories (Chen, Magoulas, & Macredie, 2004; Clewley, Chen, & Liu, 2010). It was found that there is significant interaction effect between cognitive style and search experience in terms of time spent in Web searching (Palmquist & Kim, 2000). That is, search experience contributes to a decrease in time spent retrieving information by wholistic searchers. A more detailed analysis of eye tracking data such as the time spent looking at information objects would provide more insights into cognitive processes in search interactions.

Eye Gaze and Search Interface Design

The use of eye tracking equipment for capturing searchers' fixation patterns provides a rich set of data to understand whether searchers read the document surrogates (e.g. summary and metadata) and more importantly, how searchers attend to different components of search results or search interfaces (see Lund, 2016, for a recent review). In an eye tracking study of the effect of query-based snippet length in Web searches, Cutrell and Guan (2007) found that user task performance depends on the type of search tasks (informational vs. navigational). Kules and Capra's (2012) study of the searcher's gaze behaviour in searching a faceted library catalogue showed that facets account for approximately 10–30% of interface use and the use of interface elements varies by search stages within a session.

Recent human–computer interaction and IR research has focused on users' cognitive aspects in search interactions by measuring gaze patterns, an indicator of searcher attention and cognitive processes (Dumais, Buscher, & Cutrell, 2010; Kules, Capra, Banta, & Sierra, 2009; Lorigo et al., 2008). In analysing gaze patterns, researchers have used tools to select regions of the displayed stimuli in search interfaces, also known as areas of interest (AOI) (Bojko, 2013). For example, Kemman, Kleppe, and Maarseveen's (2013) study that compared search interfaces with visible and collapsible facets reveals no significant difference in the user's AOI on the facets panel. Kules et al.'s (2009) study of user interactions with faceted search interface shows that users spent the most time looking at the search result items in which no distinction was made between title and abstract in terms of AOI. Similarly, Kammerer and Gerjets (2011) found that there is no significant effect of interface (list vs. tabular) in total fixation time on search results.

Even though the elements of search results or search interfaces are characterised in different ways for research purposes, it has been found that users pay more attention to the elements related to the contents of search results pages than elements such as search suggestions and URLs (e.g. Cutrell & Guan, 2007; Dumais et al., 2010; Kules et al., 2009). Users' attention to the snippets of Web pages increased when the length expanded (Cutrell & Guan, 2007). In Web search environments, users paid more attention to the top three, next top three and top advertisements than other regions, such as related searches on search engine results page (SERP) (Dumais et al., 2010), while few abstracts in SERPs from Google and Yahoo were viewed in query reformulation (Lorigo et al., 2008). However, the title of lower ranked items was considered more important than the snippets of higher ranked items (Savenkov, Braslavski, & Lebedev, 2011). Wang, Xie, and Lee's (2015) study found that the title and abstract of SERPs were ranked the top two elements in terms of total fixation time and fixation count.

These studies generally suggest that there is no significant difference in users' gaze on comparisons of search interface layouts. Users' attention to elements of interfaces depends on the length and quality of snippets on SERPs and displayed position of search results. As such, our study was designed to investigate how the user characteristics of domain knowledge, search experience and cognitive styles affect search processes and eye gaze, with or without the MeSH terms as feedback in different types of search interfaces for searching complex biomedical questions.

Methods

To understand domain experts' interactions with search user interfaces, a user experiment was conducted in a laboratory setting within the framework of interactive IR (Ruthven & Kelly, 2011). Eye tracking equipment was used to capture users' gaze behaviour for better understanding of user interactions. The task involved participants performing searches on clinical information for patients, with the aim of finding and saving as many relevant documents as possible. The participants were mostly students with search engine experience and some academic background in the biomedical domain. Each user was assigned eight search questions in total, with a seven-minute limit for each question. The experiment took about 90 minutes in total. The study was approved by the Science & Medical Delegated Ethics Review Committee at The Australian National University (Protocol number 2013/002)

and the Human Research Ethics Committee at Charles Sturt University (Protocol number 2013/033).

Participants

A total of 32 people participated in the study. Gender balance was achieved and most participants were students (13 undergraduate and 15 postgraduate), and young (19 were aged between 18 and 24, and 9 were between 25 and 34). The other four participants were postdoctoral researchers and academic staff. The majority (27 of 32) had not used MeSH, but most had substantial experience using general search engines, such as Google and Bing (half reported daily use and 12 reported use several times a day or more). The participants had background knowledge in the domains of biology, biotechnology, medical science, neuroscience and biomedical engineering, or some knowledge of biology in their prior learning.

Experimental Design

We used a $4 \times 4 \times 2$ factorial design with four search interfaces, controlled search question pairs and cognitive styles. A 4×4 Graeco–Latin square design was used to arrange the experimental conditions (Kirk, 2013). The sample consisted of 256 search sessions (32 participants $\times$ 8 search questions).

Search Interfaces

Participants searched on four different search user interfaces, with a single search system behind the scenes. The four search interfaces were distinguished by whether MeSH terms were presented and how the displayed MeSH terms were generated (Figure 1). Our interface design was inspired by state-of-the-art search systems, such as those used in EBSCOhost and ProQuest.

Interface 'A' mimicked Web search and other search systems with no controlled vocabulary. This interface had a brief task description at the top, a conventional search box and button and each result was represented with its title, authors, publication details and abstract where available. Full text was not available so the results were not clickable. Users judged their success on the titles and abstracts alone.

Interface 'B' (Figure 1(i)) added MeSH terms to the interface. After the user's query was run, MeSH terms from all results were collated; the 10 most frequent were displayed at the top of the screen. This mimics the per-query suggestions produced by systems like ProQuest. MeSH terms were introduced with 'Try:' and were clickable: if a user clicked a term, their query was refined to include the MeSH term and then re-run. It was hoped that the label, and the fact they work as links, would encourage users to interact with them.

Interface 'C' (Figure 1(ii)) used the same MeSH terms as 'B' but displayed them alongside each document, where they may have been more (or less) visible. It is a hybrid of interfaces 'B' and 'D'.

Interface 'D' mimicked EBSCOhost and similar systems that provide indexing terms alongside each document. As well as the standard elements from interface 'A', interface 'D' displayed the MeSH terms associated with each document, as part of that document's surrogate (Figure 1(iii)). Again, terms were introduced with 'Try:' and were clickable.

Figure 1. Three of the four search interfaces in the study. (i) Screenshot of Interface 'B', suggestions per-query and displayed at top. (ii) Screenshot of Interface 'C', suggestions per-query and displayed at top. The keywords are based on the query. (iii) Screenshot of Interface 'D', suggestions per-document and displayed with the document. The keywords are based on the retrieved documents.

Search Questions

We used search questions from a subset of the clinical topics from OHSUMED (Hersh, Buckley, Leone, & Hickam, 1994), originally created for IR system evaluation. They represented a subset of records from the MEDLINE database that the US National Library of Medicine had made available for testing purposes. The questions were slightly rewritten so they read as instructions to the participants (see Figure 2 for an example). Search questions were selected to cover a range of difficulties and were then randomly paired off to produce four pairs of questions. A final question, the same for all participants, was used for training purposes.

Procedure

Participants were given brief instructions about the search task and system features, followed by a practice search question and then the searches proper. They were informed the test collection was incomplete and out-of-date since the OHSUMED test collection (Hersh, Buckley et al., 1994) was used, with MEDLINE data from 1987 to 1991. User interaction data recorded included: all queries, mouse clicks, retrieved and saved documents, time spent and eye movements. Electroencephalogram readings were also captured, and these data will be reported elsewhere.

Background and exit questionnaires collected demographic information and asked participants about their perception of the search process. Participants' opinions of the tasks and the interfaces were sought, including the difficulty of search task, system usefulness and whether the suggested keywords were noticed or used (Interfaces B, C and D). Finally, information on participants' cognitive styles was collected by a computerised test (Peterson et al., 2003), which took a further 15 min to complete.

Hardware and Software

The search system was built on Solr, with the search results ranked by default relevance score. The MeSH terms were not specifically weighted. Eye gaze data were recorded from two Sony VFCB-EX480B infrared (IR) cameras which are controlled by Seeing Machines (https://www.seeingmachines.com) FaceLab 4.5 software (https://www.ekstremmakina.com/EKSTREM/product/facelab/index.html) and attached to a dedicated machine running Windows 7. At the same time, EyeWorks Design and EyeWorks Record (https://www.eyetracking.com/Software/EyeWorks) were used to present instructions for the corresponding search tasks during the experiment. Gaze points were recorded at 60 Hz, and the eye gaze data included the x and y coordinates of where the eye was looking on the screen, as well as the time that gaze point is recorded.

> Imagine that you are 63-year-old male with acute renal failure probably 2nd to aminoglycosides/contrast dye.
> You would like to find information about acute tubular necrosis due to aminoglycosides, contrast dye, outcome and treatment.

Figure 2. An example OHSUMED search topic, reworded for the participants.

Data Analysis

Recordings were analysed to see how often there were fixations in different parts of document surrogates (i.e. different elements of the interfaces), and therefore how often people looked at each part. Four common AOI were specified using EyeWorks Analyze: title, author, abstract and MeSH (except for Interface A, without MeSH) to investigate which elements received attention. Fixations were specified as gazes within a five-pixel radius which lasted at least 75 millisecond (ms), using the algorithms within EyeWorks Analyze (see Figure 3 for an example). The parameters are consistent with those of Rayner (1998).

We used a logarithmic cross-ratio analysis (Fleiss, Levin, & Paik, 2003; Saracevic, Kantor, Chamis, & Trivison, 1988) to determine the relationship between individual differences and their relationship to gazes. We measured participants' domain knowledge in biology as the number of undergraduate and postgraduate classes taken. The mean values (and therefore cut-points) were 11.5 and 2.2 courses, respectively. Frequency of search engine use was cut at 4.22 (4 is equivalent to daily use), whereas experience with online databases was cut at the mean of 2.5 (2 is equivalent to less than five years' experience).

The E-CSA-WA (Extended Cognitive Style Analysis–Wholistic Analytic) test produced a Wholistic Analytic ratio (WA ratio) for each participant. People tend to be wholistic when the ratio is close to 0 and a ratio closer to 2 or above indicates a tendency to be analytic. The cut-off ratio score at 1.32 (M = 1.31, SD = .24) for our analysis was consistent with previous studies (Clewley et al., 2010; Yuan, Zhang, Chen, & Avery, 2011).

Results

Search Interfaces and Gaze

Large and significant differences were observed in the attention paid to different interface elements amongst the four interfaces (Figure 4). Documents in Interfaces B, C and D received similar amounts of attention to the title, author and abstract. Searchers looked at abstracts more often than other elements of documents.
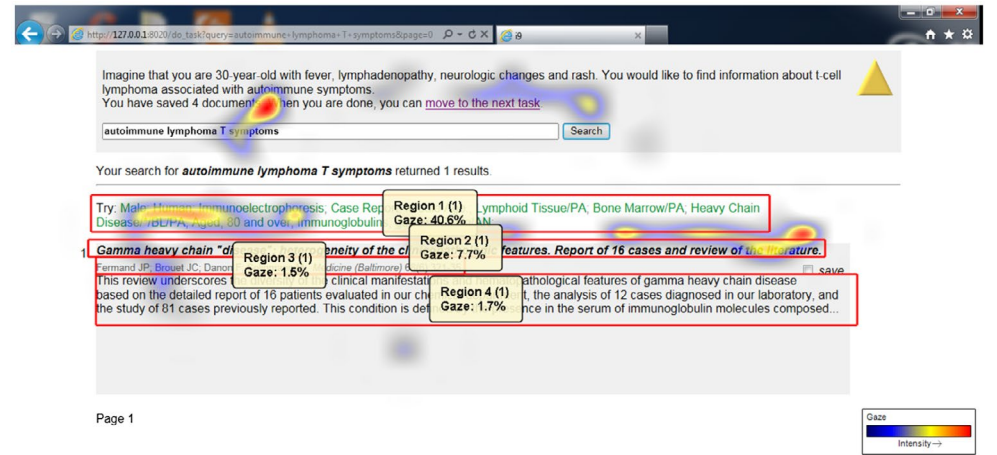
Figure 3. EyeWorks GazeSpot illustrating specification of areas of interest (AOI) and identification of gazes.
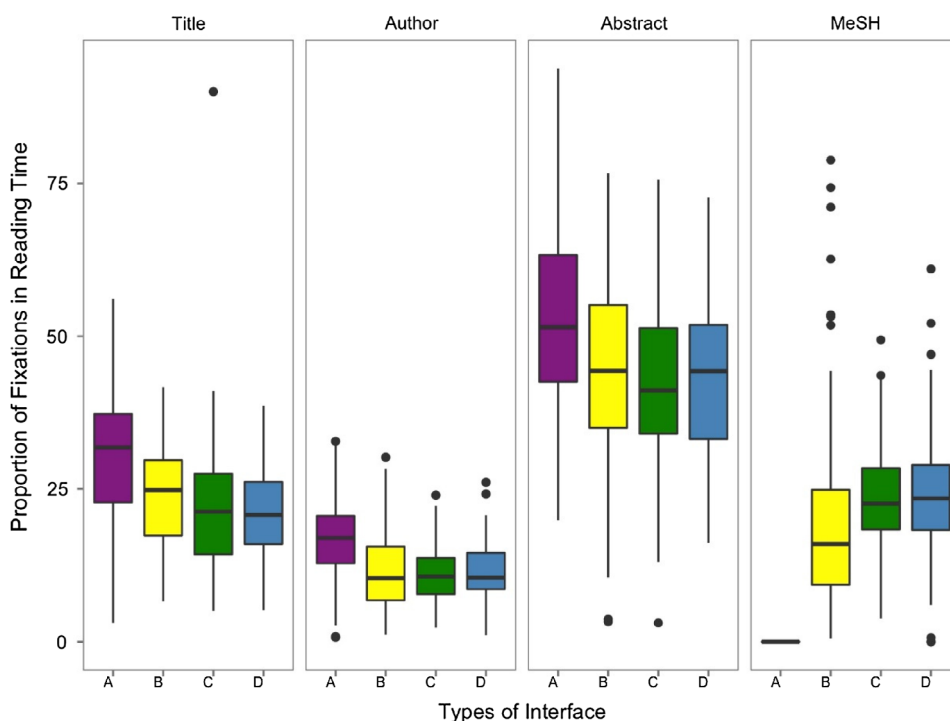
Figure 4. Fixations by area of interest, for each interface.

More specifically, it was the search interfaces that were significant in terms of the differences to the proportion of fixations on titles ($F(3, 252) = 11.05$, $p < .001$), on authors ($F(3, 252) = 10.82$, $p < .001$), on abstracts ($F(3, 252) = 6.75$, $p < .001$) and on MeSH terms ($F(3, 252) = 78.78$, $p < .001$). Tukey HSD (Honest Significant Difference), a multiple comparison procedure and *post hoc* test, indicated that differences in fixations on title, author and abstract are due to Interface A, the project's baseline search interface.

With respect to MeSH terms, significant differences were observed between Interfaces B and C, and Interfaces B and D (TukeyHSD, $p < .01$ for both comparisons). In Interface B (Figure 1(i)), the relatively low average and wide spread attention to MeSH terms may be because MeSH terms were only displayed under the search box area, whereas MeSH terms were displayed alongside each document in Interfaces C and D. These results support the hypothesis that search interfaces have significant impact on eye gaze behaviour in terms of the proportions of fixations in reading time.

Individual Differences and Gaze

Domain Knowledge

The results as reported in Table 1 suggested that there was a statistically significant relationship between the users' background knowledge in biology at the undergraduate level and their time spent looking at document titles. In other words, users with high level of undergraduate biology knowledge were less likely to pay attention to the title.

Table 1. Summary of the relationship between eye gaze and undergraduate biology background.

Areas of interest	Cut point (mean)	Odds ratio	Log odds	Stand. error	t–Value	Stat. signif.
Title	24.33	.54	−.61	.30	−2.06	Yes
Author	12.53	.78	−.25	.29	−.86	No
Abstract	45.81	1.61	.48	.29	1.65	No
MeSH	17.34	1.55	.44	.29	1.50	No

Notes: *N* undergraduate biology background = 256, *N* eye gaze = 256; statistical significance at 95%.

Table 2. Summary of the relationship between eye gaze and postgraduate biology background.

Areas of interest	Cut point (mean)	Odds ratio	Log odds	Stand. error	t–Value	Stat. signif.
Title	24.33	.45	−.81	.34	−2.40	Yes
Author	12.53	.68	−.39	.33	−1.18	No
Abstract	45.81	1.55	.44	.32	1.37	No
MeSH	17.34	1.42	.35	.32	1.09	No

Note: *N* postgraduate biology background = 256, *N* eye gaze = 256; statistical significance at 95%.

The results as reported in Table 2 indicated that there was a statistically significant relationship between the users' background knowledge in biology at the postgraduate level and their proportion of fixations in reading time of title. Users with high level of postgraduate biology knowledge were less likely to pay attention to the title.

Overall, Tables 1 and 2 revealed that users with high level of knowledge in biology at either the undergraduate or the postgraduate level were less likely to pay attention to the title in search results.

Search Experience

Statistically significant relationships were revealed between the users' frequencies of using a search engine and the proportion of fixations in reading time, as shown in Table 3. Specifically, people with high frequencies of using a search engine were more likely to pay attention to the author and MeSH terms, but were less likely to pay attention to the abstract of search results.

Further analysis of the effect of search experience, interfaces and their interactions on reading time indicated that interfaces contributed to significant differences in time spent on the title ($F(3, 250) = 9.76$, $p < .001$), author ($F(3, 250) = 11.14$, $p < .001$) and abstract ($F(3, 250) = 7.68$, $p < .001$). The effect of search engine experience on the time spent on abstract was also statistically significant ($F(2, 250) = 4.21$, $p < .05$).

However, the effect of search engine experience, interfaces and their interactions on the time spent on MeSH terms (excluding Interface A) was not significant in terms of search engine experience ($F(1, 188) = .05$, $p > .05$) and interface ($F(2, 188) = 1.55$, $p > .05$) (Fox, 2016). Because of the significant interaction effect of interfaces and search experiences ($F(2, 188) = 5.69$, $p < .01$), we examined the contrast between the three types of interfaces and the high/low levels of search experiences by the Westfall multiple testing procedure for general linear hypotheses (Hothorn, Bretz, & Westfall, 2008). It revealed that Interface B attracted much more attention in terms of proportion of fixations in MeSH reading time by low experienced searchers than high experienced searchers (Westfall, adjusted $p < .05$). There was no significant relationship between the users' online database experience and their gaze patterns.

Table 3. Summary of the relationship between eye gaze and search engine experience.

Areas of interest	Cut point (mean)	Odds ratio	Log odds	Stand. error	t–Value	Stat. signif.
Title	24.33	.80	−.22	.26	−.87	No
Author	12.53	2.32	.84	.26	3.20	Yes
Abstract	45.80	.37	−.98	.27	−3.65	Yes
MeSH	17.34	2.22	.80	.26	3.03	Yes

Note: N search engine = 256, N eye gaze = 256; statistical significance at 95%.

Table 4. Summary of the relationship between eye gaze and cognitive styles.

Areas of interest	Cut point (mean)	Odds ratio	Log odds	Stand. error	t–Value	Stat. signif.
Title	24.33	.68	−.39	.25	−1.54	No
Author	12.53	.94	−.07	.25	−.26	No
Abstract	45.81	.99	−.01	.25	−.02	No
MeSH	17.34	1.05	.05	.25	.19	No

Note: N cognitive styles = 256, N eye gaze = 256; statistical significance at 95%.
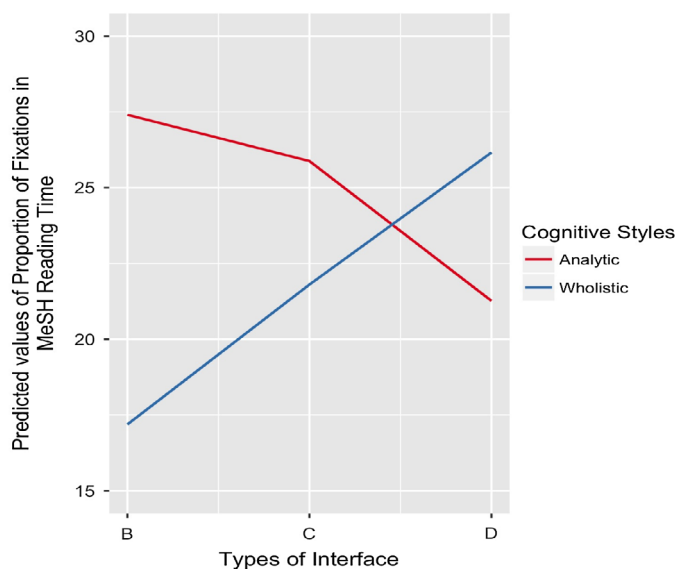
Cognitive Style
The results as reported in Table 4 suggested that there was no significant relationship between the users' cognitive style and gaze. Further analysis of the effect of cognitive styles, interfaces and their interactions on gaze indicated that interfaces made significant differences to the time spent looking at title ($F(3, 248) = 9.59$, $p < .001$), author ($F(3, 248) = 11.0$, $p < .001$) and abstract ($F(3, 248) = 7.50$, $p < .001$), but it was not the case for cognitive styles.

However, the effect of cognitive styles, interfaces and their interactions on the time looking at MeSH terms (excluding Interface A) was statistically significant in terms of cognitive style and interface interactions ($F(2, 188) = 5.31$, $p < .01$), and weakly significant in terms of cognitive style ($F(1, 188) = 2.74$, $p < .1$) (Fox, 2016). Further analysis of the contrasts between the three types of interfaces and the analytic/wholistic cognitive styles was conducted using the Westfall procedure (Hothorn et al., 2008). The results revealed that wholistic searchers' attention to MeSH terms between Interfaces B and D was significant (Westfall, adjusted $p < .05$). There was a statistically significant difference between the analytic and wholistic searchers when interacting with Interface B. These results suggest wholistic searchers preferred Interface D (mimicked EBSCOhost) to Interface B (mimicked ProQuest) in terms of the proportions of fixation time on MeSH terms. This was also reflected in the analytic searchers' preference for Interface B (see Figure 5).

Interaction effect of cognitive styles in Interface B was significant. Interaction effect of Interfaces B and D for wholistic searchers was also significant.

Summary of Individual Differences and Gaze Patterns
Table 5 provides a summary of the relationship between individual differences and gaze patterns. It suggests that users with a high level of domain knowledge were less likely to look at the title, whereas users with less experience using search engines were less likely to look at the abstract. However, users with more experience using search engines were more likely to pay attention to the author and MeSH terms. Users' online database experience or cognitive style alone did not affect their gaze patterns. Interestingly, users with low levels of search engine experience (using search engines several times a day or more) were more attracted to MeSH terms displayed in Interface B. Regarding cognitive styles, wholistic

Figure 5. Interaction plot of interface and cognitive style, in time spent looking at MeSH terms.

Table 5. Summary of the relationship between individual differences and gaze patterns.

	Domain knowledge		Search experience		
Areas of interest	Undergraduate	Postgraduate	Search engine	Online database	Cognitive style
Title	○	○	–	–	–
Author	–	–	●	–	–
Abstract	–	–	○	–	–
MeSH	–	–	●	–	–

Note: The relationship is not statistically significant (–), positively significant (●) or negatively significant (○) at 95%).

searchers were more attracted to MeSH terms in Interface D than Interface B, while analytic searchers were attracted to MeSH terms in Interface B.

Discussion and Conclusion

In this study, significant differences were found in the attention paid to different interface elements, amongst the four interfaces. Searchers looked at abstracts more often than other elements of documents. These results support the hypothesis that search interfaces have significant impact on eye gaze behaviour when users have complex questions.

These findings have confirmed the importance of users' attention to the elements related to topics (i.e. abstract of documents or snippets of SERPs) for extracting relevant information when they search to support complex questions (Kules et al., 2009; Vakkari, Luoma, & Pöntinen, 2014; Wang et al., 2015), but users rarely viewed the abstract when they do informational search tasks (Lorigo et al., 2008).

The findings overall suggest that experienced search engine users are more likely to pay attention to the elements of author and MeSH terms, but less to the abstract. Research on Web search user behaviour has revealed that users pay more attention to the elements related

to the contents of search results pages than others, such as search suggestions and URLs (Cutrell & Guan, 2007; Dumais et al., 2010; Kules et al., 2009). In this study, experienced search engine users attended to the elements of author and MeSH terms, which are usually not present in modern search engines such as Google. In particular, MeSH terms displayed in Interfaces C and D have received significantly more attention by experienced search engine users. These results indicate that search engine experience and search interfaces have a significant impact on eye gaze behaviour. So interface elements such as the MeSH terms will receive more attention if these terms are displayed alongside each document for experienced searchers.

This could be partially explained by possible exposure to search interfaces such as EBSCOhost and Ovid that provide users with the opportunity to filter search results by author and subject terms (and other elements) which are presented alongside the search results as facets. Discovery tools also enable limiting of results through the presentation of predetermined facets such as author, subjects and other elements alongside the search results. Therefore, those who reported a high frequency of using a search engine may have experienced the location of these elements alongside the results in online databases which could explain why more attention was paid to the MeSH terms when displayed alongside search results in this study.

With respect to the effect of domain knowledge on eye gaze, it was found that users with a high level of domain knowledge are less likely to attend to the element of title, as demonstrated by this study's undergraduate and postgraduate biology users with a high level of biology knowledge. However, a previous finding has suggested that titles of lower ranked items receive more attention than the snippets of higher ranked items (Savenkov et al., 2011), and the title and abstract of SERPs receive more attention than other elements by general public users (Wang et al., 2015). As such, this study has provided evidence that a user's level of domain knowledge affects eye gaze behaviour. The relative importance of elements in search interfaces and the ranked position of displayed search results need further investigation.

This study's finding that there were significant interaction effects between cognitive styles and search interfaces has supported previous research in which cognitive styles were associated with user preferences of interface design (Chen et al., 2004; Clewley et al., 2010). Specifically, wholistic users' preference of having an overview before drilling down to details was confirmed in the most amount of attention received for MeSH terms in Interface D, while analytic users' preference of looking for specific information was supported by their attention to MeSH terms in Interface B. These findings have provided more detailed understanding of the relationship between cognitive style and user interaction with search interfaces.

Implications for Practice

In view of these results and the perceived importance of MeSH, what does this mean for the library and information practitioner and how he/she should approach the decision of how to present default search options and results for library users accessing a database via a particular publisher platform (i.e. Ovid or EBSCOhost)? Does the library opt to, where possible, set the mapping to subject terms feature as a default in biomedical databases presented to searchers, given the perceived importance of MeSH terms demonstrated by this

study? Furthermore, how much detail should be provided in the results screen, given the fact that users with a high level of domain knowledge are less likely to spend a lot of time viewing title elements of search results displays and wholistic searchers spending the most amount of time spent gazing at MeSH terms? Search interfaces such as Ovid and EBSCOhost can be configured to permit varying levels of details for default search results display, in addition to the default search mode (i.e. Basic, Advanced and the ability to map to subject terms). Although users can change the default settings during a search session for Ovid and EBSCOhost, it is not known how many users change the default settings and whether the majority of users simply opt to use the default settings presented. However, these decisions may be influenced by other factors such as the type of library service, the perception of the domain knowledge levels of the majority of users and other considerations. It may be worth exploring further what factors determine a library's decision relating to the presentation of default search modes and search results display.

Disclosure Statement

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of ALIA or any author affiliated organisation.

Funding

Notes on Contributors

Ying-Hsang Liu is a lecturer in Information Management at the School of Information Studies and full member at the Institute for Land, Water and Society in Charles Sturt University. He is also an honorary lecturer at the Research School of Computer Science at the Australian National University. He holds a PhD in Information Science from the School of Communication & Information at Rutgers, The State University of New Jersey in the US.

Paul Thomas is a senior applied scientist with Microsoft. He holds a PhD in computer science from the Australian National University.

Marijana Bacic is a PhD candidate at the School of Information Studies at Charles Sturt University and recently held the position of chief librarian at Monash Health.

Tom Gedeon is a professor of Computer Science, College of Engineering and Computer Science at The Australian National University. He is head of the Information and Human Centred Computing (iHcc) research group.

Xindi Li is a master's student at the Department of Computer Science and Engineering, University of California, San Diego in the US.

ORCID

Ying-Hsang Liu http://orcid.org/0000-0001-6504-4598
Marijana Bacic http://orcid.org/0000-0001-6664-3330

References

Agosti, M., Cisco, D., Di Nunzio, G. M., Masiero, I., & Melucci, M. (2010). I-TEL-u: A query suggestion tool for integrating heterogeneous contexts in a digital library. In M. Lalmas, J. Jose, A. Rauber, F. Sebastiani, & I. Frommholz (Eds.), *Research and advanced technology for digital libraries: 14th European Conference, ECDL 2010, Glasgow, UK, September 6–10, 2010. Proceedings* (pp. 397–400). Berlin: Springer.

Behnert, C., & Lewandowski, D. (2017). A framework for designing retrieval effectiveness studies of library information systems using human relevance assessments. *Journal of Documentation, 73*, 509–527. doi:10.1108/JD-08-2016-0099

Belkin, N. J., Marchetti, P. G., & Cool, C. (1993). Braque: Design of an interface to support user interaction in information retrieval. *Information Processing & Management, 29*, 325–344. doi:10.1016/0306-4573(93)90059-M

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. New York, NY: Rosenfeld Media.

Chen, S. Y., Magoulas, G. D., & Macredie, R. D. (2004). Cognitive styles and users' responses to structured information representation. *International Journal on Digital Libraries, 4*, 93–107. doi:10.1007/s00799-003-0073-5

Clewley, N., Chen, S. Y., & Liu, X. H. (2010). Cognitive styles and search engine preferences field dependence/independence vs holism/serialism. *Journal of Documentation, 66*, 585–603. doi:10.1108/00220411011052966

Cutrell, E., & Guan, Z. (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 (CHI '07),* 407–416. doi:10.1145/1240624.1240690

Dumais, S. T., Buscher, G., & Cutrell, E. (2010). Individual differences in gaze patterns for Web search. *Proceeding of the Third Symposium on Information Interaction in Context (IIiX '10),* 185–194. doi:10.1145/1840784.1840812

Fenichel, C. H. (1981). Online searching: Measures that discriminate among users with different types of experiences. *Journal of the American Society for Information Science, 32*, 23–32. doi:10.1002/Asi.4630320104

Fleiss, J. L., Levin, B. A., & Paik, M. C. (2003). *Statistical methods for rates and proportions*. Hoboken, NJ: Wiley.

Ford, N. (2015). *Introduction to information behaviour*. London: Facet.

Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). Thousand Oaks, CA: Sage.

Golub, K., Soergel, D., Buchanan, G., Tudhope, D., Lykke, M., & Hiom, D. (2016). A framework for evaluating automatic indexing or classification in the context of retrieval. *Journal of the Association for Information Science & Technology, 67*, 3–16. doi:10.1002/asi.23600

Hearst, M. A. (1995). TileBars: Visualization of term distribution information in full text information access. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CH '95),* 59–66. doi:10.1145/223904.223912

Hearst, M. A. (2009). *Search user interfaces*. New York, NY: Cambridge University Press.

Hearst, M. A. (2011). 'Natural' search user interfaces. *Communications of the ACM, 54*, 60–67. doi:10.1145/2018396.2018414

Heinstrom, J. (2010). *From fear to flow: Personality and information interaction*. Oxford: Chandos.

Hersh, W., Buckley, C., Leone, T. J., & Hickam, D. (1994). OHSUMED: An interactive retrieval evaluation and new large test collection for research. *Proceedings of the ACM SIGIR Conference (SIGIR '94),* 192–201.

Hersh, W. R., Haynes, R. B., & McKibbon, K. A. (1994). A performance and failure analysis of saphire with a MEDLINE test collection. *Journal of the American Medical Informatics Association, 1*, 51–60. doi:10.1136/jamia.1994.95236136

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346–363. doi:10.1002/bimj.200810425

Howard, H. (1982). Measures that discriminate among online searchers with different training and experience. *Online Review, 6*, 315–327. doi:10.1108/eb024106

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

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing & Management, 36*, 207–227. doi:10.1016/S0306-4573(99)00056-4

Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., & Gay, G. (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. *ACM Transactions on Information Systems, 25*(2). doi:10.1145/1229179.1229181

Kammerer, Y., & Gerjets, P. (2011). Effects of search interface and internet-specific epistemic beliefs on source evaluations during Web search for medical information: An eye-tracking study. *Behaviour & Information Technology, 31*, 83–97. doi:10.1080/0144929X.2011.599040

Kemman, M., Kleppe, M., & Maarseveen, J. (2013). Eye tracking the use of a collapsible facets panel in a search interface. In T. Aalberg, C. Papatheodorou, M. Dobreva, G. Tsakonas, & C. J. Farrugia (Eds.), *Research and advanced technology for digital libraries: International conference on theory and practice of digital libraries, TPDL 2013, Valletta, Malta, September 22–26, 2013* (pp. 405–408). Berlin: Springer.

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.

Koenemann, J., & Belkin, N. J. (1996). A case for interaction: A study of interactive information retrieval behavior and effectiveness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96),* 205–212. doi:10.1145/238386.238487

Kules, B., & Capra, R. (2012). Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. *Journal of the American Society for Information Science and Technology, 63*, 114–138. doi:10.1002/asi.21647

Kules, B., Capra, R., Banta, M., & Sierra, T. (2009). What do exploratory searchers look at in a faceted search interface? *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '09),* 313–322. doi:10.1145/1555400.1555452

Leroy, G., Xu, J., Chung, W., Eggers, S., & Chen, H. (2007). An end user evaluation of query formulation and results review tools in three medical meta-search engines. *International Journal of Medical Informatics, 76*, 780–789. doi:10.1016/j.ijmedinf.2006.08.001

Liu, Y.-H. (2010). A meta-analysis of the effects of search experience on search performance in terms of the recall measure in controlled IR user experiments. *Proceedings of the Fifteenth Australasian Document Computing Symposium (ADCS '10)* (pp. 105–110). Melbourne: School of Computer Science and IT, RMIT University.

Liu, Y.-H., & Belkin, N. J. (2008). Query reformulation, search performance, and term suggestion devices in question-answering tasks. *Proceedings of the Second International Symposium on Information Interaction in Context (IIiX '08),* 21–26. doi:10.1145/1414694.1414702

Liu, Y.-H., & Wacholder, N. (2017). Evaluating the impact of MeSH (Medical Subject Headings) terms on different types of searchers. *Information Processing & Management, 53*, 851–870. doi:10.1016/j.ipm.2017.03.004

Lorigo, L., Haridasan, M., Brynjarsdottir, H., Xia, L., Joachims, T., Gay, G., … Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology, 59*, 1041–1052. doi:10.1002/asi.20794

Lund, H. (2016). Eye tracking in library and information science: A literature review. *Library Hi Tech, 34*, 585–614. doi:10.1108/LHT-07-2016-0085

Lykke, M., Price, S., & Delcambre, L. (2012). How doctors search: A study of query behaviour and the impact on search results. *Information Processing & Management, 48*, 1151–1170. doi:10.1016/j.ipm.2012.02.006

McKibbon, K. A., & Fridsma, D. B. (2006). Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. *Journal of the American Medical Informatics Association, 13*, 653–659. doi:10.1197/jamia.M2087

Mu, X., Lu, K., & Ryu, H. (2014). Explicitly integrating MeSH thesaurus help into health information retrieval systems: An empirical user study. *Information Processing & Management, 50*, 24–40. doi:10.1016/j.ipm.2013.03.005

Niu, X., & Kelly, D. (2014). The use of query suggestions during information search. *Information Processing & Management, 50*, 218–234. doi:10.1016/J.Ipm.2013.09.002

Palmquist, R. A., & Kim, K. S. (2000). Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science, 51*, 558–566. doi:10.1002/(SICI)1097-4571(2000)51:6<558::AID-ASI7>3.0.CO;2-9

Pao, M. L., Grefsheim, S. F., Barclay, M. L., Woolliscroft, J. O., McQuillan, M., & Shipman, B. L. (1993). Factors affecting students use of MEDLINE. *Computers and Biomedical Research, 26*, 541–555. doi:10.1006/Cbmr.1993.1038

Peterson, E. R., Deary, I. J., & Austin, E. J. (2003). The reliability of riding's cognitive style analysis test. *Personality and Individual Differences, 34*, 881–891.

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

Ruthven, I., & Kelly, D. (Eds.). (2011). *Interactive information seeking, behaviour and retrieval.* London: Facet.

Saracevic, T., Kantor, P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science, 39*, 161–176.

Savenkov, D., Braslavski, P., & Lebedev, M. (2011). Search snippet evaluation at Yandex: Lessons learned and future directions. In P. Forner, J. Gonzalo, J. Kekäläinen, M. Lalmas, & M. de Rijke (Eds.), *Lecture notes in computer science* (Vol. 6941, pp. 14–25). Berlin: Springer.

Shiri, A. (2012). *Powering search: The role of thesauri in new information environments.* Medford, NJ: Information Today.

Swan, R. C., & Allan, J. (1998). Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems. *Proceedings of the ACM SIGIR Conference (SIGIR '98), 173*–181. doi:10.1145/290941.290987

Tamine, L., & Chouquet, C. (2017). On the impact of domain expertise on query formulation, relevance assessment and retrieval performance in clinical settings. *Information Processing & Management, 53*, 332–350. doi:10.1016/j.ipm.2016.11.004

Vakkari, P., Luoma, A., & Pöntinen, J. (2014). Books' interest grading and dwell time in metadata in selecting fiction. *Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14), 28*–37. doi:10.1145/2637002.2637008

Wacholder, N. (2011). Interactive query formulation. *Annual Review of Information Science and Technology, 45*, 157–196. doi:10.1002/aris.2011.1440450111

Wang, Y., Xie, I., & Lee, S. (2015). Explore eye movement patterns in search result evaluation and individual document evaluation. *Proceedings of the Association for Information Science and Technology, 52*(1), 1–4. doi:10.1002/pra2.2015.1450520100144

White, R. W., & Marchionini, G. (2007). Examining the effectiveness of real-time query expansion. *Information Processing & Management, 43*, 685–704. doi:10.1016/j.ipm.2006.06.005

Wilson, M. L. (2011). Search user interface design. *Synthesis Lectures on Information Concepts, Retrieval, and Services, 3*(3), 1–143. doi:10.2200/S00371ED1V01Y201111ICR020

Yuan, X. J., Zhang, X. M., Chen, C. M., & Avery, J. M. (2011). Seeking information with an information visualization system: A study of cognitive styles. *Information Research, 16*(4). Retrieved from https://InformationR.net/ir/16-4/paper499.html