Only Forward? Toward Understanding Human Visual Behaviour when Examining Search Results

George Buchanan
University of Melbourne
Parkville, Victoria, 3010
dana.mckay@unimelb.edu.au

Dana McKay
University of Melbourne,
Parkville, Victoria, 3010
george.buchanan@unimelb.edu.au

Eduardo Velloso
University of Melbourne
Parkville, Victoria, 3010
eduardo.velloso@unimelb.edu.au

Alistair Moffat
University of Melbourne
Parkville, Victoria, 3010
ammoffat@unimelb.edu.au

Andrew Turpin
University of Melbourne
Parkville, Victoria, 3010
aturpin@unimelb.edu.au

Falk Scholer
RMIT University
124 La Trobe St., Melbourne, Vic 3000
falk.scholer@rmit.edu.au

ABSTRACT

Search is near-ubiquitous in human society, being used for entertainment, health, financial and business information seeking. Traditional methods of search evaluation have assumed that searchers move forward through search results in a linear manner; early eye tracking studies have suggested the same. Recent research, though, including eye-tracking data, has demonstrated a number of counter-cases, particularly for search in complex conditions. These examples highlight how little we understand where humans look or what they are doing when examining search results: in this paper we undertake a broad survey of the literature to refine this research question, and suggest avenues for developing a model of behaviour.

CCS CONCEPTS

• **Human-centered computing~HCI theory, concepts and models** • **Information systems~Search interfaces** • Information systems~Presentation of retrieval results

KEYWORDS

Search, human vision, information seeking, information retrieval evaluation, eyetracking.

OzCHI '17, Nov. 28-Dec. 1, 2017, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-5379-3/17/11…$15.00
https://doi.org/10.1145/3152771.3156165

ACM Reference format:

Dana McKay, et al. Only Forward? Toward Understanding Human Visual Behaviour when Examining Search Results. In *Proceedings of the 29th Australian Conference on Human-Computer Interaction, Brisbane, QLD, Australia, November 2017 (OzCHI 2017)*, 6 pages. https://doi.org/10.1145/3152771.3156165

1 INTRODUCTION

Google alone processes 3.5 billion queries per day, or 1.2 trillion queries per year [1]. When the queries on other public search sites, e.g. Baidu, Yahoo or Bing are added, the total tops 6.6 billion [2]. This is more than one query per day for every literate human over the age of 14[3]. When we consider searches on other services—local file systems, library services, enterprise search, etc. [1]—it is fair to say that search is a ubiquitous human activity. Given the ubiquity of search, our understanding of search behaviour is almost bizarre in its simplicity.

If you ask users how they look at a page of internet search results, they will likely tell you they start at the top, hoping the first one will answer their requirements, and—if their need is not met—work down vertically and systematically until they find one that does [2]. This same assumption—that users' search result evaluation is systematic and vertical, looking for only a single result—is at the heart of infinite scroll. If a searcher gets 'to the bottom' of search results the design assumption underpinning infinite scroll is that they have not found what they are looking for, rather than that they are attempting e.g. collection understanding. Similarly, in the models that underpin traditional information retrieval evaluation methods, it is assumed that users will move down a strictly ranked list of search results without ever moving back up. Early eyetracking studies of search supported this assumption, with abundant gaze time spent on the first search results in a set and a direct inverse correlation between result rank and overall gaze time [3].

---

1 http://www.internetlivestats.com/google-search-statistics/
2 http://www.smartinsights.com/search-engine-marketing/search-engine-statistics/
3 http://data.worldbank.org/indicator/

There are other domains in which 'only-forward' movement has been assumed and found lacking. Marshall and Bly demonstrated that physical reading is not the linear activity we might expect [4]; McKay demonstrated the same for digital reading [5]. Buchanan and Loizides [6] demonstrated that readers flip back and forth while selecting journal articles, even though they spend more time on the first page in aggregate. Ebook triage is similarly typified by significant flicking back and forth with more time spent in the earlier parts of the book [7].

Researchers have long understood that a single search result is not enough to meet all information needs; users understand this too [8, 9]. It is only relatively recently, however, that we have seen anything to challenge the only-forward assumption about search result evaluation. MacFarlane et al. [10] have demonstrated that people with dyslexia move to the bottom of the page and work up a list of search results, largely due to differences in short term memory. This means that for people with dyslexia, infinite scroll is an accessibility nightmare, and that—while they too may spend more time looking at the first search result—an only-forward assumption is plain wrong. Similarly, Thomas et al. [11] demonstrated that the difficulty of a search task affects the extent to which users engage with the result list, and that reading is far from linear and monotonic: like document and ebook triage it is characterised by backward-and-forward movements. Searchers engaging in simple tasks were more likely to view only a few documents, whereas searchers with more complex information needs examined the results to deeper levels. The same study showed that the choice of documents already viewed influenced what searchers viewed next: they did not simply click down a list in order [11].

Given the ubiquity of search, it is perhaps surprising that so little is known about how searchers actually view and move between search results. In this paper we present a case for more closely investigating how searchers examine results from a visual perspective, and suggest means by which this examination might be achieved. We first briefly describe the most relevant existing literatures on search and vision, then survey the history of search evaluation methods. We discuss previous work on search result evaluation, then finally describe the gaps in the literature, pointing to some avenues for addressing them.

2 INTERACTION WITH SEARCH RESULTS

The received wisdom on searcher query construction has long been that it is naïve. Searchers' short, simple queries and their use of defaults [12, 13] have are seen as symptoms of naïveté, not evidence of a sophisticated effort/reward trade off with search systems. Recent literature suggests that in fact user strategies are highly adapted, and that searchers both change defaults and user longer queries when it benefits them to do so [14].

There is a similar received wisdom about search result viewing: searchers 'only click on the first result' and 'only look at one page of results'. As with the clichés about query construction, this set of assumptions has been based on a range of research (e.g. [12],[15]). Recent work has provided many clues that searchers' behaviour is more complex and sophisticated than we have typically allowed [16]. This section addresses the types of complexities that have been found in studies of vision and search result presentation, as a result of task type or complexity, and due to individual differences between users.

Traditional user models of interactive information retrieval presume that searchers explore result pages linearly, starting from the top and working their way down the retrieved results [17]. This is also a key underlying assumption of most widely-used IR evaluation metrics [18, 19].

2.1 Presentation and Vision Effects

During document selection, a high proportion of user time is spent reading both search result pages and candidate documents [20]. In both cases we might assume linear reading, but the truth of this assumption is open to question.

As far back as the 60's, researchers have known that people do not read every word in a page when looking for information. Saracevic reported that when searching for specific material, readers looked at very few elements, such as the title and conclusion of an academic paper, rather than thoroughly reading its entire content [21]. These findings were corroborated by recent studies that examined these behaviours both in print and online. Buchanan and Loizides studied document triage, i.e. the initial encounter with a document where the reader assesses the relevance of the document in relation to their information need [6]. They found that regardless of the length of the document, most readers spent over half of the time on the first page.

These non-linear patterns of document reading are mirrored when search engine result pages (SERPs) are viewed. Eye tracking studies confirm that users' attention is not evenly spread across the page [3, 22, 23]. The average first-fixation time is, unsurprisingly, inversely correlated with the document's rank in the result list when averaged over many users [22]. On average, users' gaze is mostly directed to the first 3 or 4 search results and the further from the top the item is, the less attention it receives. Further, if users do not see their desired result in the top eight results, they prefer to generate a new query rather than move to the second page of results.

Thus far, the linear account appears to be intact. However, there are numerous counter-points. One well-established phenomenon is the increased likelihood of links at the very bottom of the first search result page being clicked, compared to those immediately above them [24]. The increasing availability of eyetracking equipment has given researchers the opportunity to examine this: clicks at the bottom of pages appear to follow a user seeing the end of the page, and choosing to click a result rather than navigating to the next result page.

Eyetracking also revealed previously unseen phenomena: at any moment in time, the user's attention is localised even within a typical result page of ten items. Thomas et al. found that during search, users maintain a region of interest a mere two or three summaries wide, continuously switching between them [11, 24]. In fact, searchers seldom consider any result independently of its immediate neighbours. Again, linearity proves only a partial account of actual behaviour.

The parts of a document summary are another area where attention is not evenly distributed. These summaries typically have three parts: the title, the URL, and the snippet. Previous research has studied the effect of varying these features, changing the length of the snippet or highlighting query terms in bold [22, 25, 26]. The impact of these changes, has, however, only been measured through clickthroughs on single summaries. Despite being a suitable candidate for eye tracking evaluation, to our knowledge there are no studies that measure the direct impact of these factors impact on our viewing behaviour Furthermore, it has been recently shown that users change their search tactics when the interface to a collection is altered [14]. Such changes are happening at a larger scale, as Google and other internet search engines include images, ads, and Wikipedia article summaries in place of the former simple result list.

2.2 Task Effects

Task complexity affects search result behaviour. While we might expect users to use navigation buttons to view search results and candidate documents in browser tab, this behaviour changes in complex task conditions. Blandford et al [27] observed users' search sessions in person, and identified strategies such as using multiple tabs in more complex tasks. They also noted that concurrent search tasks affect each other, causing mutual interruptions and digression. Behaviour is different again for non-text search tasks, for example image search [28].

Our insight into different tasks continues to develop: e.g. Spina et al. [29] showed substantial differences between those looking for jobs, and those looking for potential employees on the same site. In complex tasks, searchers regularly formulate more queries, consult more documents per query and navigate across many result pages. In simple tasks, searchers often stop within the first result page of results, even if unsuccessful [30].

One key division of search tasks breaks them into three types: informational, navigational and transactional [31]. Studies have shown that these lead to different behaviours, in terms of query formulation, result inspection and the reading of candidate document. Guan and Cutrell [3] report that in navigational search, users expect to discover a match at very close to the top of the list, while in informational searches—seldom answered in one document—users consider more results and abandon a search less hastily. Liu et al, [32] showed that other aspects of search behaviour such as decision time, task completion time, and attention also vary by task type. Within a given task type, the complexity of the task leads to different behaviours [20]. Kelly et al. revealed that the cognitive complexity of the task also strongly influences searchers' visual behaviour, both across result sets and within summaries [33]. Another predictable influence on behaviour is subject expertise: expert users adapt more sophisticated strategies when needed, and domain knowledge results in different scanning behaviours [30]. In a particular search, task effects appear to be a much stronger determiner of action than overall averages of 'normal' behaviour.

2.3 Individual Differences

User behaviour, from query formulation to scanpaths, varies between users in ways that are systemic, not idiosyncratic. Studies show that the-eyetracked scanpaths of individual users over a list of results found that they vary by factors such as gender, the items the user has already inspected, and the order in which those documents were viewed [34, 35]. Aggregated user behaviour appears to match simple a only-forward narrative, but there is clear variability according to task type and complexity,

MacFarlane et al. [10] investigated the impact of dyslexia on a user's visual analysis of search results. This revealed that many dyslexic users scan from the bottom rather than the top of the visual page. Given the prevalence of dyslexia, this is a significant challenge to the top-down emphasis of previous research. Blind users also show specific patterns, viewing fewer items in a more linear pattern [36], and when other users interact through the same audio interface, it is likely this pattern will recur.

2.4 Summary

User behaviour shows a number of predictable patterns that diverge from the only-forward assumptions that were prevalent fifty years ago: users change their behaviour according to task type, task complexity, individual differences and search result presentation. We now summarise the methods that might best be used to investigate such changes.

3 METHODS FOR EVALUATING IR SYSTEMS

Traditional methods for information retrieval evaluation have not involved humans as part of the search process: the so-called Cranfield method was developed in the 1960s and consists of three core components: a set of representative search queries, a set of documents to search over, and for each query-document pair, a human-generated judgment of whether the document is "about" the same thing as the query [37]. These methods have underpinned the evaluation approach prevalent in the TREC conference and many papers appearing at SIGIR. The founding principle of these methods is that results should be ranked in strict relevance order, so that the hypothetical users of any system evaluated in this way would see the most relevant result first. In principle, this is a clearly convincing manner for presenting documents against an information need.

One critique of these methods from within the community that uses them is the heterogeneity of relevance—we cannot say, for a given user in a given context which results will be the most relevant. More relevant for the HCI community, though, is the series of works comparing the results of Cranfield evaluations with user rankings of search systems: perhaps unsurprisingly there is considerable divergence (e.g. [38, 39]).

The traditional reason given within the IR community for not evaluating systems using actual searchers is that such evaluations are (relative to automated methods) time consuming and expensive; of course, they are also demonstrably necessary.

Moving away from automated methods but without actively recruiting users is search log analysis; this has been used for years to evaluate library catalogues, digital libraries and online search systems. These studies are the ones that have led us to understand that users type in short queries and rarely change defaults (a result that in itself has recently been challenged), and that they click most often on the first result and rarely move beyond the first page. While these studies provide a solid 'birds eye view' of behaviour, they do not allow us to examine the subtleties of users' contexts nor understand their motivations.

Diary studies are another at-a-distance method that play almost to the opposite strengths of log analyses: where log analyses are effective—capturing query terms and user clicks—diary studies are poor. In contrast where log analyses fail—understanding user context and motivations—diary studies are very effective. An early example of this is Warwick et al. [40] who demonstrated that people consider more search results when there is pressure to get the correct answer.

Another at-a-distance method is the one used by Khoo and Hall [2]: they asked students to heuristically evaluate library systems then mined the resulting papers for attitudes to search in general and Google in particular. While this is not a strategy that could be readily used for a new search system, it did generate some interesting insights into search perception. It showed for example that users perceived Google as not providing 'too many' search results while the library systems they were evaluating did. This is in stark contrast to the actual number of results—the highly relevant search results generated by Google mitigate against the information overload that could result from the vastness of their number.

Remote user testing, the final at-a-distance method is typically used allow testers access to participants they may otherwise not have, and to test in a slightly more naturalistic setting. Greifeneder [41] tested this as a method and discovered that distraction is a big influencer on remote testing, so this method should be used with caution.

The in-person user testing that is contrasted with 'in absentia' methods (i.e. the Cranfield approach) in traditional IR comes in many forms [42], including formal usability testing, observational usability testing, and eye tracking studies. Seminal examples of formal usability testing of search, where there are predetermined hypotheses and a highly-controlled environment include Spool's revelation that users enter more words in a larger search box. Observational user testing has proven very rich; from Marchionini's early work [15] describing how people look for information at all through Blandford's extensive work (e.g. [43]) showing that users perform differently under complex conditions right up to Vakkari's recent work demonstrating user relevance judgements based on their search moves [44], this type of study has generated a wide range of insights into search.

Finally, and most relevant to the work presented here, are eyetracking studies. Historically technology in this space was costly and had low accuracy; it also excluded anyone who normally wears glasses as a participant. Early aggregate studies confirmed that users look more at results higher in a relevance ranked list [45]; later studies have demonstrated the situation to be far more nuanced, particularly for complex tasks [11].

While there is much we do not know about user search behaviour, one of the richer avenues for future investigation is eyegaze evaluation. We have already demonstrated that many of our early assumptions about how searchers examine results are wrong, particularly in complex situations, however we do not understand what users *are* doing in those cases. The recent increase in accuracy and reduction in cost of eyetracking systems means that for the first time it is possible to conduct extensive analysis of gaze during search result evaluation, potentially revealing much about human information seeking.

4 WHAT *ARE* WE LOOKING AT?

It is abundantly clear that users do not move only-forward through search results, despite the assumptions underpinning traditional IR evaluation. There are numerous influences on this behaviour, including task type and complexity, search results presentation and individual differences. We are only beginning to understand the individual impact of each influence on user behaviour; there is certainly no work on how they interact.

It is clear that traditional models of information retrieval evaluation can provide no further insight into these problems: we need actual users interacting with search results in a range of contexts to understand what is going on. While there are many tools we could use to understand these issues, we argue that searcher gaze is one of the least-understood pieces of the puzzle.

Search result presentation has become more complex in very recent time. Eyetracking has at the same time become readily available and relatively inexpensive. While we could observe users to understand their task behaviour, this would not afford the fine detail of eyetracking data. Understanding what users look at, how long they look at it for, and in what order allows us to understand their interactions in a way that merely observing their behaviour cannot, for example without eyetracking users of dyslexia the major differences in their behaviour from the non-dyslexic population would be imperceptible [10].

While eyetracking has given researchers a closer insight into search interaction, there remain gaps in our methods and the available data. Most eye-tracking analysis has examined total fixation time on features such as succeeding search results [45], or different parts of the detail given for each search result [25]. Scant attention has been paid to the scanpath between elements, and we lack data on how the user's previous path influences their future movements. The relationship of viewing patterns across tabs and navigation steps (e.g. going 'back' a page) is also sparsely investigated. Given the significance of these strategies in observed sessions, this is a major gap in our knowledge.

We believe that—rather than a single, general pattern of visual search behaviour over search results—there are many patterns, and that those will be influenced by task type, task complexity, search result presentation and individual differences. The interactions between these features will naturally influence behaviour. What we do not know—and cannot guess—is what these patterns would look like. A series of eyetracking studies investigating these dimensions could provide an initial insight into these behaviours, with the potential for generating a user behaviour model. This model could, in turn, be used to improve search result presentation usability and searcher effectiveness, and to predict searcher behaviour on the basis of search result presentation. It remains an open research field what that model might look like; we intend to begin investigating it.

REFERENCES

[1] Buchanan, G., Cunningham, S. J., Blandford, A., Rimmer, J. and Warwick, C. 2005. Information seeking by humanities scholars. In *Proceedings of the 9th European conference on Research and Advanced Technology for Digital Libraries* (Vienna, Austria, 2005). Springer-Verlag, 218-229. DOI:http://dx.doi.org/10.1007/11551362_20

[2] Khoo, M. and Hall, C. *What Would 'Google' Do? Users' Mental Models of a Digital Library Search Engine*. Springer Berlin Heidelberg, City, 2012.

[3] Guan, Z. and Cutrell, E. 2007. An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, 2007). ACM, 417-420. DOI:http://dx.doi.org/10.1145/1240624.1240691

[4] Marshall, C. C. and Bly, S. 2005. Turning the page on navigation. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries* (Denver, CO, USA, 2005). ACM, 225-234. DOI:http://dx.doi.org/10.1145/1065385.1065438

[5] McKay, D. 2011. A jump to the left (and then a step to the right): reading practices within academic ebooks. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (Canberra, Australia, 2011). ACM, 202-210. DOI:http://dx.doi.org/10.1145/2071536.2071569

[6] Buchanan, G. and Loizides, F. 2007. Investigating document triage on paper and electronic media. In *European Conference on Research and Advanced Technology for Digital Libraries* (Budapest, Hungary, 2007). Springer-Verlag, 416-427

[7] Hinze, A., McKay, D., Vanderschantz, N., Timpany, C. and Cunningham, S. J. 2012. Book selection behavior in the physical library: implications for ebook collections. In *ACM/IEEE-CS Joint Conference on Digital Libraries* (Washington, DC, USA, 2012). ACM, 305-314. DOI:http://dx.doi.org/10.1145/2232817.2232874

[8] Makri, S., Blandford, A. and Cox, A. L. 2008. Investigating the information-seeking behaviour of academic lawyers: From Ellis's model to design. *Inf. Process. Manage.*, 44, 2 (2008), 613-634.

[9] Attfield, S., Makri, S., Kalbach, J., Blandford, A., Gabrielle, S. and Edwards, M. 2008. Prioritisation, Resources and Search Terms: A Study of Decision-Making at the Virtual Reference Desk. In *Proceedings of the 12th European conference on Research and Advanced Technology for Digital Libraries* (Aarhus, Denmark, 2008). Springer-Verlag, 106-116. DOI:http://dx.doi.org/10.1007/978-3-540-87599-4_12

[10] MacFarlane, A., Albrair, A., Marshall, C. R. and Buchanan, G. 2012. Phonological working memory impacts on information searching: an investigation of dyslexia. In *Proceedings of the 4th Information Interaction in Context Symposium* (Nijmegen, The Netherlands, 2012). ACM, 27-34. DOI:http://dx.doi.org/10.1145/2362724.2362734

[11] Thomas, P., Scholer, F. and Moffat, A. *What Users Do: The Eyes Have It*. Springer Berlin Heidelberg, City, 2013.

[12] Siegfried, S., Bates, M. J. and Wilde, D. N. 1993. A profile of end-user searching behavior by humanities scholars: The Getty Online Searching Project Report No. 2. *Journal of the American Society for Information Science*, 44, 5 (1993), 273-291.

[13] Spink, A., Wolfram, D., Jansen, M. B. J. and Saracevic, T. 2001. Searching the Web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.*, 52, 3 (2001), 226-234.

[14] McKay, D. and Buchanan, G. 2013. Boxing clever: how searchers use and adapt to a one-box library search. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration* (Adelaide, Australia, 2013). ACM, 497-506. DOI:http://dx.doi.org/10.1145/2541016.2541031

[15] Marchionini, G. *Information seeking in electronic environments*. Cambridge University Press, 1995.

[16] Bailey, P., Moffat, A., Scholer, F. and Thomas, P. 2015. User Variability and IR System Evaluation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile, 2015). ACM, 625-634. DOI:http://dx.doi.org/10.1145/2766462.2767728

[17] Joachims, T., Granka, L., Pan, B., Hembrooke, H. and Gay, G. 2005. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (Salvador, Brazil, 2005). ACM, 154-161. DOI:http://dx.doi.org/10.1145/1076034.1076063

[18] Clarke, C. L. A., Craswell, N., Soboroff, I. and Ashkan, A. 2011. A comparative analysis of cascade measures for novelty and diversity. In *Proceedings of the fourth ACM international conference on Web search and data mining* (Hong Kong, China, 2011). ACM, 75-84. DOI:http://dx.doi.org/10.1145/1935826.1935847

[19] Moffat, A. and Zobel, J. 2008. Rank-biased precision for measurement of retrieval effectiveness. *ACM Trans. Inf. Syst.*, 27, 1 (2008), 1-27.

[20] Buscher, G., White, R. W., Dumais, S. and Huang, J. 2012. Large-scale analysis of individual and task differences in search result page examination strategies. In *Proceedings of the fifth ACM international conference on Web search and data mining* (Seattle, Washington, USA, 2012). ACM, 373-382. DOI:http://dx.doi.org/10.1145/2124295.2124341

[21] Saracevic, T. 1969. Comparative effects of titles, abstracts and full texts on relevance judgments. *Proceedings of the American Society for Information Science*, 6, 1 (1969), 293-299.

[22] Cutrell, E. and Guan, Z. 2007. What are you looking for?: an eye-tracking study of information usage in web search. In *SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, 2007). ACM, 407-416. DOI:http://dx.doi.org/10.1145/1240624.1240690

[23] Nielsen, J. and Pernice, K. *Eyetracking Web Usability*. New Riders Publishing, 2009.

[24] Jiang, J., He, D. and Allan, J. 2014. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (Gold Coast, Queensland, Australia, 2014). ACM, 607-616. DOI:http://dx.doi.org/10.1145/2600428.2609633

[25] Clarke, C. L. A., Agichtein, E., Dumais, S. and White, R. W. 2007. The influence of caption features on clickthrough patterns in web search. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (Amsterdam, The Netherlands, 2007). ACM, 135-142. DOI:http://dx.doi.org/10.1145/1277741.1277767

[26] Turpin, A., Scholer, F. and Von Billerbeck, B. *Examining the pseudo-standard web search engine results page*. Queensland University of Techology, Faculty of Information Technology, City, 2006.

[27] Blandford, A., Keith, S., Connell, I. and Edwards, H. 2004. Analytical usability evaluation for digital libraries: a case study. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries* (Tuscon, AZ, USA, 2004). ACM, 27-36. DOI:http://dx.doi.org/10.1145/996350.996360

[28] Xie, X., Liu, Y., Wang, X., Wang, M., Wu, Z., Wu, Y., Zhang, M. and Ma, S. 2017. Investigating Examination Behavior of Image Search Users. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan, 2017). ACM, 275-284. DOI:http://dx.doi.org/10.1145/3077136.3080799

[29] Spina, D., Maistro, M., Ren, Y., Sadeghi, S., Wong, W., Baldwin, T., Cavedon, L., Moffat, A., Sanderson, M., Scholer, F. and Zobel, J. 2017. Understanding User Behavior in Job and Talent Search: An Initial Investigation. In *SIGIR Workshop on eCommerce* (Tokyo, Japan, 2017)

[30] White, R. W., Dumais, S. T. and Teevan, J. 2009. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (Barcelona, Spain, 2009). ACM, 132-141. DOI:http://dx.doi.org/10.1145/1498759.1498819

[31] Broder, A. 2002. A taxonomy of web search. *SIGIR Forum*, 36, 2 (2002), 3-10.

[32] Liu, J., Cole, M. J., Liu, C., Bierig, R., Gwizdka, J., Belkin, N. J., Zhang, J. and Zhang, X. 2010. Search behaviors in different task types. In *Proceedings of the 10th annual joint conference on Digital libraries* (Gold Coast, Queensland, Australia, 2010). ACM, 69-78. DOI:http://dx.doi.org/10.1145/1816123.1816134

[33] Kelly, D., Arguello, J., Edwards, A. and Wu, W.-c. 2015. Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval* (Northampton, Massachusetts, USA, 2015). ACM, 101-110. DOI:http://dx.doi.org/10.1145/2808194.2809465

[34] Buscher, G., Dengel, A., Biedert, R. and Elst, L. V. 2012. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Trans. Interact. Intell. Syst.*, 1, 2 (2012), 1-30.

[35] Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K. and Newman, J. K. 2004. The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of the 2004 symposium on Eye tracking research \& applications* (San Antonio, Texas, 2004). ACM, 147-154. DOI:http://dx.doi.org/10.1145/968363.968391

[36] Sahib, N. G., Tombros, A. and Stockman, T. 2012. A comparative analysis of the information-seeking behavior of visually impaired and sighted searchers. *Journal of the American Society for Information Science and Technology*, 63, 2 (2012), 377-391.

[37] Sanderson, M. 2010. Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends® in Information Retrieval*, 4, 4 (2010), 247-375.

[38] Turpin, A. and Scholer, F. 2006. User performance versus precision measures for simple search tasks. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (Seattle, Washington, USA, 2006). ACM, 11-18. DOI:http://dx.doi.org/10.1145/1148170.1148176

[39] Turpin, A. H. and Hersh, W. 2001. Why batch and user evaluations do not give the same results. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval* (New Orleans, Louisiana, USA, 2001). ACM, 225-231. DOI:http://dx.doi.org/10.1145/383952.383992

[40] Warwick, C., Rimmer, J., Blandford, A., Gow, J. and Buchanan, G. 2009. Cognitive economy and satisficing in information seeking: A longitudinal study of undergraduate information behavior. *J. Am. Soc. Inf. Sci. Technol.*, 60, 12 (2009), 2402-2415.

[41] Greifeneder, E. 2011. The impact of distraction in natural environments on user experience research. In *Proceedings of the 15th international conference on Theory and practice of digital libraries: research and advanced technology for digital libraries* (Berlin, Germany, 2011). Springer-Verlag, 308-315
[42] Kelly, D. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval*, 3, 1–2 (2009), 1-224.
[43] Blandford, A., Stelmaszewska, H. and Bryan-Kinns, N. 2001. Use of multiple digital libraries: a case study. In *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries* (Roanoke, Virginia, USA, 2001). ACM, 179-188. DOI:http://dx.doi.org/10.1145/379437.379479
[44] Vakkari, P., P, J., #246 and ntinen. 2015. Result List Actions in Fiction Search. In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries* (Knoxville, Tennessee, USA, 2015). ACM, 7-16. DOI:http://dx.doi.org/10.1145/2756406.2756911
[45] Granka, L. A., Joachims, T. and Gay, G. 2004. Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (Sheffield, United Kingdom, 2004). ACM, 478-479. DOI:http://dx.doi.org/10.1145/1008992.1009079