Using Information Scent to Understand Mobile and Desktop Web Search Behavior

Kevin Ong
School of Science
RMIT University
kevin.ong@rmit.edu.au

Kalervo Järvelin
School of Information Science
University of Tampere
kalervo.jarvelin@uta.fi

Mark Sanderson
School of Science
RMIT University
mark.sanderson@rmit.edu.au

Falk Scholer
School of Science
RMIT University
falk.scholer@rmit.edu.au

ABSTRACT

This paper investigates if Information Foraging Theory can be used to understand differences in user behavior when searching on mobile and desktop web search systems. Two groups of thirty-six participants were recruited to carry out six identical web search tasks on desktop or on mobile. The search tasks were prepared with a different number and distribution of relevant documents on the first result page. Search behaviors on mobile and desktop were measurably different. Desktop participants viewed and clicked on more results but saved fewer as relevant, compared to mobile participants, when information scent level increased. Mobile participants achieved higher search accuracy than desktop participants for tasks with increasing numbers of relevant search results. Conversely, desktop participants were more accurate than mobile participants for tasks with an equal number of relevant results that were more distributed across the results page. Overall, both an increased number and better positioning of relevant search results improved the ability of participants to locate relevant results on both desktop and mobile. Participants spent more time and issued more queries on desktop, but abandoned less and saved more results for initial queries on mobile.

CCS CONCEPTS

•Human-centered computing → HCI theory, concepts and models; Text input; •Information systems → *Information retrieval query processing; Users and interactive retrieval;*

KEYWORDS

Information Foraging Theory; Search Process; Search Stopping

ACM Reference format:
Kevin Ong, Kalervo Järvelin, Mark Sanderson, and Falk Scholer. 2017. Using Information Scent to Understand Mobile and Desktop Web Search Behavior. In *Proceedings of SIGIR '17, Shinjuku, Tokyo, Japan, August 07-11, 2017,* 10 pages.

SIGIR '17, Shinjuku, Tokyo, Japan

DOI: http://dx.doi.org/10.1145/3077136.3080817

DOI: http://dx.doi.org/10.1145/3077136.3080817

1 INTRODUCTION

Information Foraging Theory [23] (IFT) seeks to understand how information seekers behave when searching. The theory compares information seeking behavior to food-foraging strategies used by animals. It posits that information seekers will adapt their behavior and gravitate towards an equilibrium that optimizes valuable information gain per unit cost [23].

An Information Scent model was proposed [6], which is a prediction model based on IFT. It suggests that information seekers will use visual cues to guide them towards relevant information sources. Such cues can come from the contents of a Search Engine Result Page (SERP), which contains information (e.g. title, URL, summary) about retrieved documents. Searchers can then make use of this information to help them decide if a document is relevant and if they will click on it. Researchers applied the Information Scent model to a study of desktop web search behavior by varying the number of relevant search results (*level*) and their distribution (*pattern*) [30]. They found that both features are predictive of some user behaviors. Searchers are more likely to abandon their search if: 1) fewer relevant search results are presented or 2) the relevant search results are in lower positions on a SERP.

It has been argued that the continual growth of mobile search has brought a paradigm shift in web search behavior. Searching on mobile and desktop can be considered as searching in different environments. Mobile is different from desktop in terms of timely access to information and differences in screen sizes [9]. The applicability of desktop-based interface research findings to mobile environments is not clear. It is therefore worth investigating whether different environments affect search behavior differently. Two comparative user studies are conducted to understand the applicability of IFT on search behavior in both mobile and desktop environments.

We focus our efforts on understanding how differences between mobile and desktop affect search behavior. We investigate the effect of staying above the fold, a concept borrowed from print-newspaper terminology, on search behavior. *Above the fold* refers to the portion of the SERP that is immediately seen on screen; *below the fold* refers to the portion that needs to be scrolled to. Using IFT, we seek to understand the extent to which information scent may influence search behavior in different environments. We address the following research questions:

RQ1: To what degree can mobile and desktop web search behavior be explained by Information Scent Level (ISL)?
We vary the number of relevant information items in a SERP and measure searchers' behavior in mobile and desktop environments.

RQ2: To what degree can mobile and desktop web search behavior be explained by Information Scent Pattern (ISP)?
We vary the distribution of a fixed number of relevant search results in a SERP and measure searchers' behavior.

RQ3: How does search behavior differ as a result of different environments?
Information visibility can influence search behavior [13]. We use different environments (desktop and mobile) as representatives of different folds to measure the differences in search behavior. We seek to understand how different levels of visibility may influence search behavior when users are given the same search tasks with identical ISL/ISP conditions, but in different environments.

2 LITERATURE REVIEW

The development of search models is studied widely by the information science community [2, 23, 30]. In this section, we discuss research related to understanding search behavior.

2.1 Web Search Behavior

A wide range of observational studies have had been conducted on web search behavior on desktop [2, 8, 10, 13] and mobile [14, 15, 18, 20, 21, 25].

Desktop Web Search: *Granka et al.* [10] studied thirty-six users focusing on their actions before the selection of the first retrieved document. They found that the users tended to focus on URLs in particular, and the first and second search results in the SERP. *Joachims et al.* [13] found that users clicked on the first result regardless of the quality of subsequent results. They observed that users tended to perform a top-down search pattern and placed substantial trust in the search engine's ordering of documents. They also observed that the quality of retrieved results influenced clicking behavior. When the SERPs were made deliberately worse, users clicked on fewer relevant search results. *Cutrell and Guan* [8] studied twenty-two participants using an eye tracker while they conducted *informational* and *navigational* tasks [3]. They observed that users preferred longer snippets for informational tasks and shorter for navigational tasks. User's search accuracy (i.e. clicking on relevant search results) was improved for informational tasks but degraded for navigational tasks when snippet lengths were increased. Similar to previous work [13], the researchers ascertained that the ranking of relevant search results influenced user behavior. When relevant results were placed in lower positions in a SERP, users were less likely to locate those results. *Azzopardi et al.* [2] studied thirty-six undergraduate students. They associated *query cost* with the degree of difficulty in issuing search queries. An inverse relationship between query number and search depth was observed. They found that when search interfaces got more complicated, users issued fewer queries and increased search depth. *Maxwell et al.* [22] later proposed six search stopping strategies based on *disgust* and *frustration* point rules, to predict the moment when a user would stop searching. One strategy, stoppage after a certain fixed depth, was found to be accurate.

Searching on Mobile: Search behavior on mobile can be different from desktop [7, 9, 14, 15, 20]. *Jones et al.* [14] studied twenty computer science students and staff on two tasks using desktop screens and mobile (simulated) screens. Mobile participants were twice as likely to fail in finding relevant information and twice as likely to use the search functions, compared to desktop participants. Searchers would rather use the search function than attempt to locate the relevant information manually, when it was harder to find on the page. However, it was noted that both groups were using actual physical keyboards, which might influence their preference for search functions for the mobile participants. Finding relevant information involves entering queries and examining results. When input was unhindered, search increased [14]. When given actual devices, however, searchers issued both shorter and fewer queries on mobile than on desktop [15]. In a study by *Kamvar and Baluja* [16], the average number of queries per session on mobile was two. A later comparative mobile study by *Song et al.* [27] found that the average length of users' issued queries increased but this was attributed to a better auto-completion feature on mobile. It was also observed that the number of query submissions per session on mobile were smaller than on desktop. *Ghose et al.* [9] observed that the ranking effects of results were greater on mobile than on desktop due to the limited number of results that can be displayed at once. Scrolling through more results incurs cognitive costs, as the searcher has to remember past results. *Lagun et al.* [20] studied mobile search behavior of thirty participants. Similar to past work [13], they observed that position bias affected user search accuracy when searching on mobile devices. However, they found users spent more time on second and third results compared to the first. *Ren et al.* [25] examined mobile search behavior in a large indoor retail space by analyzing ISP logs over a one year period and found that mobile Web searching and browsing behavior was different. *Church et al.* [7] carried out a diary study over four weeks to study twenty users' mobile information needs. They found that mobile information needs differ significantly from general Web (i.e. desktop) needs. As users increasingly use mobile as their only device for search[1], mobile information needs and search deserve further attention.

2.2 Search Strategies

Considering search strategies, *Klöckner et al.* [19] observed two distinct approaches: breadth-first (skimming through a number of snippets first before clicking) and depth-first search (clicking each document sequentially before looking at new snippets). They observed that users who preferred depth-first search were significantly more likely to click a promising link before looking at others within the list. *Teevan et al.* [28] interviewed fifteen Computer Science graduates twice a day over five days, grouping them into *filers* (people who organized information using fixed structures) and

[1]https://storage.googleapis.com/think/docs/twg-how-people-use-their-devices-2016.pdf

pilers (people who maintained unstructured information organization). They observed that filers and pilers relied on two different search strategies. Filers relied more on keyword searches, while pilers were more likely to use site search engines (such as eBay site search) rather than generic search engines. *Aula et al.* [1] used an eye tracker to study twenty-eight users and also observed two types of search strategy patterns: *economic* and *exhaustive*. Similar to depth-first searchers [19], they found that *economic* users examined results sequentially from the top-down and clicked on the first relevant search result they saw, whereas *exhaustive* searchers examined all results before even considering which to click. *White and Drucker* [29] studied the extent of users' search behavior variability over a five month period. They concluded that information seekers can be classified into two broad categories: *Navigators* and *Explorers*. Navigators, like filers, employ a search strategy to organize information, with directed searches and topical coherence in the search trails. Explorers, similar to pilers, have information overlap (re-visits to multiple links) when searching for information. *Kim et al.* [17] investigated search examination strategies on different screen sizes with thirty-two participants using *Klöckner et al. [19]*'s taxonomy. They observed that users implemented more breadth-first and fewer depth-first strategies on a large screen than on a small screen, contrary to *Klöckner et al. [19]*'s findings. These previous works looked at search strategies on desktop and suggested that user factors and individual differences resulted in two distinct search strategies of interaction with search engines. *Li et al.* [21] discussed the concept of *good abandonment*. It was considered as good abandonment when a user's information need was already satisfied by information displayed on the SERP itself resulting in no result clicks. The good abandonment rate was found to be significantly higher on mobile than on desktop. In general, the ease of query inputs and the difficulty in finding relevant information would both encourage additional reformulations beyond the first queries.

2.3 Information Foraging

Information Foraging Theory (IFT) was proposed by *Pirolli and Card* [23] to understand web search behavior from an ecological standpoint. Information seekers, analogous to food foraging animals, will evolve over time to optimize their information seeking, gathering, and consumption behaviors. There are three derivative models from IFT: Diet Selection (factors that determine one type of information over another), Information Patch (factors to remain within sources of information) and Information Scent (factors that determine value of information based on visual cues and metadata). The use of Information Scent [5] has been suggested to explain a user's web search behavior on SERPs [8, 30].

Card et al. [4] later developed the Web Behavior Graphs methodology using IFT, to illustrate search structures performed by users. They concluded that Information Scent played an important role in the methodology. *Cutrell and Guan* [8] found that positions of relevant search results influenced searcher's behavior and suggested the use of IFT for future work. *Wu et al.* [30] then conducted an IFT based study to understand user behavior on desktop. SERPs with different *levels* and *distributions* of Information Scent conditions were prepared. Participants viewed documents in lower positions when more relevant search results were present. They also abandoned their search earlier if relevant search results were only shown later on the SERPs. A cognitive scale, Need For Cognition (NFC) measures the extent to which a person enjoys tasks that requires thinking. *Wu et al.* found that for users interacting with SERPs with a medium level of information scent, search behavior was different depending on a user's NFC; users with higher NFC tended to ignore lower-ranked search results and to paginate less.

Past work has demonstrated the differences between desktop and mobile search behavior. Additionally, it has been shown that IFT can be used to understand search behavior better. However, we found no comparative work that discussed the influence of different environments on search behavior or search strategies using IFT.

3 EXPERIMENTAL SETUP

The experimental design is based on previous work by *Wu et al. [30]*, where users are asked to carry out searches with an experimental IR system modeled closely on a web search engine. Users are asked to mark the search results which they believe to be relevant for the current search topic. Our study has some modifications and one key difference. Instead of being required to view each search document and indicate the item as relevant, participants in the experiments save each result as relevant using a checkbox displayed next to each snippet directly on the search results page. We made it optional for participants to view actual documents. This is done so as to estimate the likelihood that users would only view a snippet to decide if a document is relevant. This is particularly important for mobile, due to higher good abandonment rate [21]. The study was run in two environments: searching on a desktop and on a mobile device. For both, participants were asked to find relevant search results for a provided task until they were satisfied. Six open ended search topics and one demo topic were prepared beforehand for the user studies. The first search result page was fixed in content and layout so as to ensure particular levels and patterns of relevant and non-relevant documents were present in the SERP. Figure 1 shows the layout of retrieved documents for each of the ISL (Low, Medium, and High) and ISP (Bursting, Persistent, and Disrupted). ISL-Low (ILL), ISL-Medium (ILM), ISL-High (ILH) contained one, three and five relevant search results from first position respectively. ISP-Bursting (IPB), ISP-Persistent (IPP), ISP-Disrupted (IPD) distributed four relevant search results on the first SERP. ISP showed zero, half and all relevant search results above the fold under IPB, IPP and IPD conditions on mobile.

3.1 Participants

Seventy-two students from various disciplines, aged between 18 to 47, were recruited in a local campus library to participate in the user studies via opportunistic sampling. The study was reviewed and approved by the RMIT University Human Research Ethics Committee. All participants claimed to be English language and search engine proficient. They completed a total of 429 search tasks on both desktop and mobile, with our custom built search engine. We excluded 3 search tasks from 2 participants due to problems with logging and system stability issues.

	ISL			ISP		
Rank	ILL	ILM	ILH	IPB	IPP	IPD
1	**R**	**R**	**R**	–	**R**	**R**
2	–	**R**	**R**	–	**R**	**R**
3	–	**R**	**R**	–	–	**R**
Above the fold (mobile) ↑ Below the fold (mobile) ↓						
4	–	–	**R**	**R**	–	**R**
5	–	–	**R**	**R**	**R**	–
6	–	–	–	**R**	–	–
7	–	–	–	**R**	–	–
8	–	–	–	–	**R**	–
Above the fold (desktop) ↑ Below the fold (desktop) ↓						
9	–	–	–	–	–	–
10	–	–	–	–	–	–

Figure 1: SERP display following first query for each ISL/ISP condition. (R = relevant; – = not relevant).

3.2 Tasks

Participants were divided into two groups of thirty-six, to carry out their searches using either desktop or mobile devices. Each participant completed the same six search tasks. Half the tasks to investigate the influence of ISL and the remaining half on the influence of ISP. These were the same six informational tasks developed by *Wu et al.* [30]. Each task was presented to participants with a predefined topic description and participants were free to express their query as they saw fit. However, for their first query for each task, the participants saw a predefined SERP drawn from one of the ISL/ISP conditions in Figure 1. Topics and Information Scent conditions were rotated and counter-balanced to avoid possible learning and ordering effects. Therefore, each task with identical ISL/ISP conditions was seen twelve times (half on desktops and half on mobiles) across all the participants, but in a random order.

All videos, images, maps, PDFs, and related links were removed so that all tasks showed the same text search results. All result pages for the first query for each topic were cached locally, and documents were shown should the participants chose to open any link. The topics were chosen to be relatively simple, which should take no more than 5–7 minutes to complete. Participants were told not to spend more than 45 minutes in total, but could freely allocate their time between topics. They were also free to leave the study at any time - though none did. At the end of the session, they were compensated with a $20 voucher for their participation.

Procedure. All participants were first introduced to the experiment and were asked to fill out a pre-task questionnaire on their search experience and expertise. They then performed a simple test to collect information on their typing behavior and were given the demo task for them to familiarize with the search interface, as well as to reinforce the perception that the search results were live. The search interface was created to have a similar feel to a commercial search engine (see Figure 2).

After reading each topic motivation and description, participants were free to type in any query into the interface and were asked to find as many relevant search results as possible until they were satisfied. Participants could save relevant results at any time by marking a checkbox next to each result in a SERP. After their first query for each task, the search results for subsequent reformulations were retrieved from a commercial search engine. We did not prepare the SERPs for additional reformulations according to ISL/ISP conditions because reformulation search behavior is different from initial search behavior [26].

Table 1: Average Relevant Scent (ARS) values.

	ILL	ILM	ILH	IPB	IPP	IPD
ARS value	1.0	2.0	3.0	5.5	4.0	2.5

SERP Construction. A set of relevant and non-relevant search results were constructed by issuing queries to a commercial search engine. We used the top issued queries from previous work [30] and submitted our own non-relevant search queries. We combined the relevant and non-relevant search results into a SERP according to the order dictated by Figure 1. Three assessors then evaluated the search result lists based on the topic statement. Results that were not agreed upon by all three assessors were discarded until enough search results were gathered to construct the SERP pages for all six topics. We also placed three relevant search results in the twelfth, fifteenth, and eighteenth positions on the second SERP. This was displayed to participants who choose to view results beyond the first ten search results, for all six result list patterns, so that participants would not find viewing the second page to be fruitless.

3.3 Apparatus

Desktop. Participants in this group completed the search tasks on a laptop with a 15" screen. We gathered information about their preferred device as the keyboard may not be the one they are familiar with. However, we found no correlation in regard to keyboard familiarity and typing behavior by the time they finished the demo task. Participants were also provided with a mouse to interact with the search results. However, they could choose to use the track pad if they preferred. In the desktop environment, eight results are visible above the fold.

Mobile. Participants in this group could choose to use either an iPhone 6 or Samsung S6 to complete their task. The iPhone 6 display is 4.7" while the Samsung S6 display is 5.1". To account for the differences in screen size, the font sizes on both devices were calibrated as closely as possible to ensure that the number of characters across both screens were similar when viewing the SERP. Three results were visible above the fold on both devices.

3.4 Measurements

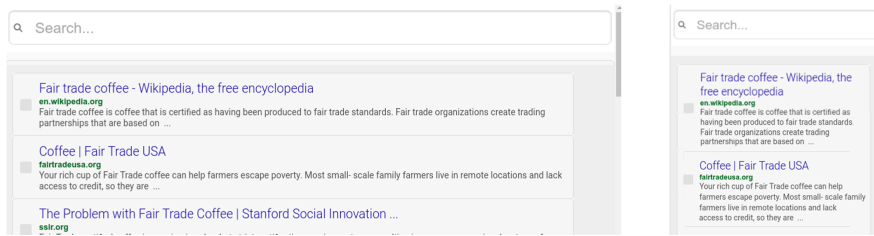
We record two types of search behavior: task level and initial query level.

Task level search behavior:

- **TimeTotal**: Total Time spent examining search results per task.
- **NumQuery**: Number of query submissions per task.

Initial query search behavior:

- **QueryAction**: The first action carried out after an *initial* query submission, apart from viewing/marking documents



(a) desktop (b) mobile

Figure 2: Mockups of the search Interface used by participants for desktop and mobile search respectively

on the SERPs: (1) issuing a new query (*Reformulation* action), (2) viewing the second SERP (*Pagination* action) without reformulation or (3) ending the task after viewing the first SERP (*Stopping* action) without (1) and (2).

- **Time**: Time spent examining search results for the first query per task.
- **NumPage**: Number of SERP paginations per task.
- **NumClick**: Number of documents examined for each search result set.
- **DRC**: Lowest search result position among all clicked documents, 0 if no results were clicked.
- **DRV**: Lowest search result position that became visible on screen during a search, logged using a Javascript package. If the participant fetches a second page, search depth ranges from 11–20.
- **ROA**: Rate Of Abandonment, the rate of not clicking or saving any document as relevant after the initial query submission for each task.
- **DRS**: Lowest position of search results on the first SERP saved as relevant per participant.
- **STotal**: Total number of search results on the first SERP saved as relevant per topic per participant.
- **SRele**: Total number of relevant search results on the first SERP (according to ISL/ISP relevance conditions) saved as relevant per topic per participant.
- **SRele%**: The percentage of saved relevant search results against the total number of relevant search results on the first SERP. Higher value indicates increased success rate.

3.5 Calculating Average Relevant Scent (ARS)

We define Average Relevant Scent (ARS) as the average rank position of relevant documents on the first SERP:

$$ARS = \frac{\sum pos_d}{\sum doc} \quad (1)$$

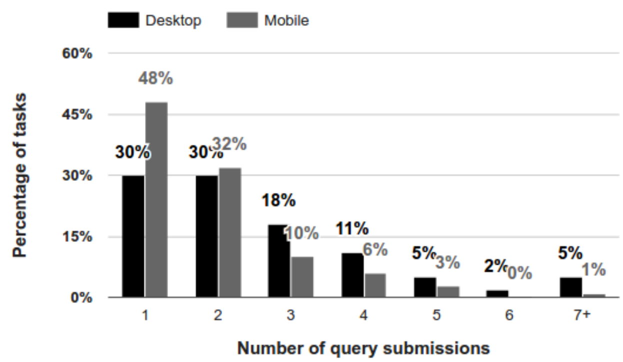


Figure 3: Number of query submissions per task.

where pos_d is the position of relevant document d on the first SERP and doc is the number of relevant documents on the first SERP. ARS attempts to summarize the depth to which participants are willing to examine documents on the first SERP based on the different information scent conditions. These values are given in Table 1.

4 RESULTS

Across the two groups of 36 participants and 429 search tasks, 414 mobile and 568 desktop queries were submitted over the study. The mean time taken for each task was less than 5 minutes. Shortest and longest time per task on desktop was 81 seconds and 18.5 minutes; on mobile, 71 seconds and 15.4 minutes.

Figure 3 illustrates the distribution of query submissions across all the tasks. 60% and 80% of the tasks were completed with 1 to 2 query submissions on desktop and mobile environments respectively. The distribution of queries was more gradual on desktop than mobile. Only 12% of desktop and 4% of mobile tasks exceeded 4 queries per topic.

Table 2: Search behavior at the environment level.

	TimeTotal (sec)	Time	NumQuery	NumPage	NumClick	DRC	DRV	ROA	DRS	STotal	SRele
Desktop	267****	64	2.67****	.46*	.58*	2.54*	14.15****	22%*	3.11*	1.81	1.70
Mobile	214****	61	1.92****	.35*	.42*	1.59*	11.81****	14%*	3.49*	1.94	1.77

Wilcoxon signed-rank test: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$

Table 3: Search behavior measures (M, SD) by Information Scent Level (ISL).

ISL	Desktop			Mobile		
Measures	**ILL**	**ILM**	**ILH**	ILL	ILM	ILH
TimeTotal#	265.58 (110.60)*	**280.20 (185.70)**	260.17 (143.99)	201.77 (95.04)*	214.26 (111.84)	**218.63 (146.65)**
Time#	47.92 (28.19)	57.37 (38.11)	**62.94 (39.85)**	43.06 (21.20)	**62.37 (44.21)**	53.91 (26.39)
NumQuery	**2.97 (2.17)***	2.74 (1.74)	2.94 (2.39)**	2.03 (1.06)*	**2.22 (2.13)**	1.89 (1.43)**
NumPage	.31 (.47)	.46 (.51)	**.49 (.51)**	.28 (.45)	**.31 (.47)**	**.31 (.47)**
NumClick	.39 (.49)	.40 (.81)	**.69 (1.16)***	.19 (.40)	**.56 (.91)**	.31 (.92)*
DRC	1.53 (3.88)****	**2.17 (5.08)******	2.00 (4.35)*	.50 (1.84)****	.75 (1.16)****	**1.31 (3.90)***
DRV	12.44 (5.18)*	14.17 (5.38)**	**14.34 (5.41)****	10.83 (5.60)*	11.17 (5.74)**	**11.69 (5.19)****
ROA[1]	39%	23%	**17%**	28%	**11%**	**11%**
DRS	.53 (.56)***	1.86 (1.44)	**3.17 (2.55)**	1.42 (2.13)***	2.42 (1.65)	**3.69 (2.45)**
STotal	.53 (.56)*	1.60 (1.33)	**2.37 (1.88)**	.92 (.91)*	1.86 (1.25)	**2.78 (1.74)**
SRele	.50 (.51)	1.49 (1.09)	**2.26 (1.80)**	.61 (.49)	1.64 (1.07)	**2.61 (1.55)**
SRele%	**50%**	49%	45%	**61%**	55%	52%

Mean and (standard deviation) values are shown. Significant differences are indicated for same ISL conditions across different environments.

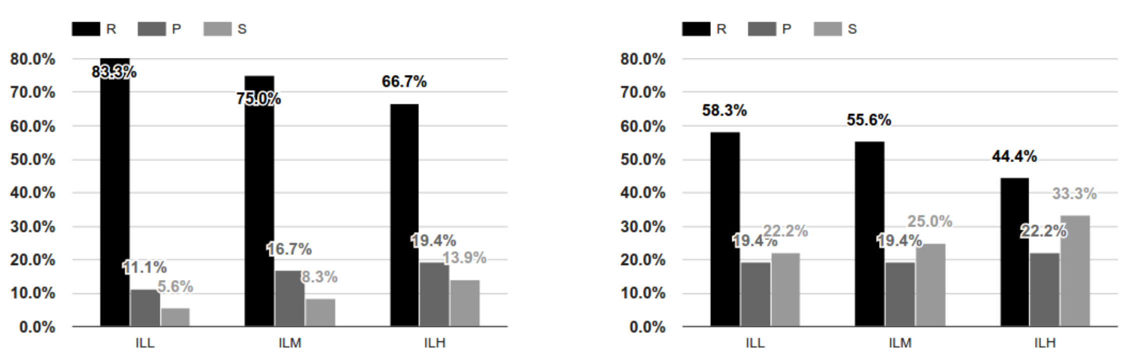
[1] Lowest value for ROA were bold for higher user interaction.

- indicates Student's t-test, otherwise Chi-squared test. Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$.

Desktop Versus Mobile Search Behavior: General search behavior trends are reported in Table 2. The Wilcoxon signed-rank test is used to evaluate the significance of differences in distributions of values between the two environments, desktop and mobile. We report significant differences between both environments where $p < 0.05$. Participants spent significantly longer TimeTotal per task on desktops compared to mobiles ($p < .0001$). The participants submitted 2.67 and 1.92 queries on average, for desktop and mobile respectively ($p < .0001$). Participants on desktop issued more queries (NumQuery) and viewed lower rank positions (DRV) than on mobile ($p < .0001$). In addition, NumPage, NumClick, and DRC were significantly different between mobile and desktop ($p < .05$). DRS was lower ($p < .05$) on mobile for the first query. Overall, search behavior across desktops and mobiles was measurably different. While desktop participants searched and viewed more results, fewer results were saved for their first queries.

ISL & Search Behavior: Considering the influence of different Information Scent Level (ISL) conditions between desktop and mobile search behavior, Figure 4 shows the distribution of three main QueryActions across tasks for different ISL conditions. On desktop, *R*eformulation decreased by 20% from 83.3% to 66.7% while both *P*agination and *S*topping increased by 75% and 148% respectively, when ISL increased from ILL to ILH. On mobile, *R* decreased by 24% from 58.3% to 44.4% while both *P* and *S* increased by 14% and 50% respectively, when ISL increased from ILL to ILH conditions. Between ILL and ILM condition on mobile, *P* did not increase.

We test the significance of changes in search behavior due to ISL using the Chi-square test, with results reported in Table 3. Significant differences between different ISL conditions for $p < 0.05$ are reported. The critical values for X^2 across the ISL conditions for both desktop and mobile environments are reported separately in Table 5 (left side). There are significant differences between different ISL conditions for both desktop and mobile for: DRS, STotal and SRele ($p < .0001$). For mobile, NumClick ($p < .05$) and DRC ($p < .001$) were significantly different between conditions. These differences indicated that ISL manipulations influenced search behavior in different environments.



(a) Desktop

(b) Mobile

Figure 4: Distribution of Search Behavior by QueryAction: Reformulation (R), Pagination (P) and Stopping (S) for the first query controlled by ISL conditions on both desktop and mobile.

Significant differences across different ISL conditions are reported for desktop, followed by mobile. The highest values, within the same environments, are denoted in bold. In the *Desktop* columns of Table 3, STotal (X^2 = 39.20, p < .0001) and SRele (X^2 = 37.62, p < .0001) increased with higher ISL. DRS are deeper with increasing ISL conditions (X^2 = 65.11, p < .0001). SRele% dropped by 10% from 50% to 45%, as ISL increased. We also observed that ROA reduced by 51% as ISL increased from ILL to ILH, from 39% to 17%. The deepest document click-through rate increased by 38% from ILL to ILM before dropping 8% in the ILH condition.

The *Mobile* columns of Table 3 show that participants clicked on documents in lower positions (X^2 = 14.37, p < .001) as ISL increased. Similar to desktop, both STotal (X^2 = 33.67, p < .0001) and SRele (X^2 = 44.45, p < .0001) register lower values under higher ISL conditions. Participants also tended to save documents in lower positions (DRS) with increased ISL (X^2 = 37.40, p < .0001). SRele% dropped 15% from 61% to 52% as the information scent increased. Time spent under ILM condition were 42% and 15% more, compared to the ILL and ILH conditions respectively (X^2 = 113.19, p < .001). NumClick was highest under ILM condition (X^2 = 7.00, p < .05). Therefore, higher ISL did not always contribute to higher NumClick. There was also no differences between the ILM and ILH conditions for ROA and NumPage. Therefore, search behavior measures did not consistently increase between ILM and ILH under mobile ISL conditions. In general, the measures for search behavior increased while ROA decreased as the ISL increased from Low to High on desktop but not on mobile.

ISP & Search Behavior: Next, the impact of different ISP conditions on search behavior is considered. Figure 5 shows the three QueryActions for desktop and mobile. For the ISP conditions on desktop, there was no consistent observable trend for the QueryActions. The Reformulation rate was consistently above 60%, while the Pagination rate dipped by 30% to 19.4% in the IPP condition before rising back to 27.8% under the IPD condition. Search stopping behavior on desktop showed a 34% increase from 8.3% to 11.1% from the IPB condition to both the IPP and IPD conditions. For mobile, search stopping behavior increased by 83% from 16.7% to 30.6% when the average positions of relevant search results moved into higher positions. Apart from Search Stopping behavior on mobile which increased when ARS improved, desktop and mobile ISP conditions did not show a consistent trend.

Changes in performance for search behavior are reported in Table 4 for ISP conditions. The critical values for X^2 across the ISP conditions within a single environment (desktop or mobile) are reported separately in Table 5 (right side). DRS between the different conditions within desktop (X^2 = 17.12, p < .001) and mobile (X^2 = 25.08, p < .0001) are significantly different. Time is different between desktop ISP conditions (X^2 = 82.34, p < .0001) and TimeTotal is different between mobile ISP conditions (X^2 = 14.01, p < .001). This indicates that, in general, the ISP manipulations do not heavily influence search behavior across the different ISP conditions in either environment.

In the *Desktop* columns of Table 4, document depth corresponded to the change in ISP conditions, DRS decreased from 5.06 to 3.17 as ISP changed from IPB to IPD conditions (X^2 = 17.12, p < .001). Time and NumClick also increased as ARS moved from 5.5 to 2.5 (IPB to IPD conditions). ROA also decreased by 50% from the IPB to IPD conditions. We observed that both STotal, SRele increased as relevant results are displayed earlier on the SERPs. Search accuracy improved by 22% from 45% to 55% as the ARS changed from 5.5 to 2.5. Overall, participants saved documents in higher positions and spent more time when the relevant search results were placed in higher positions.

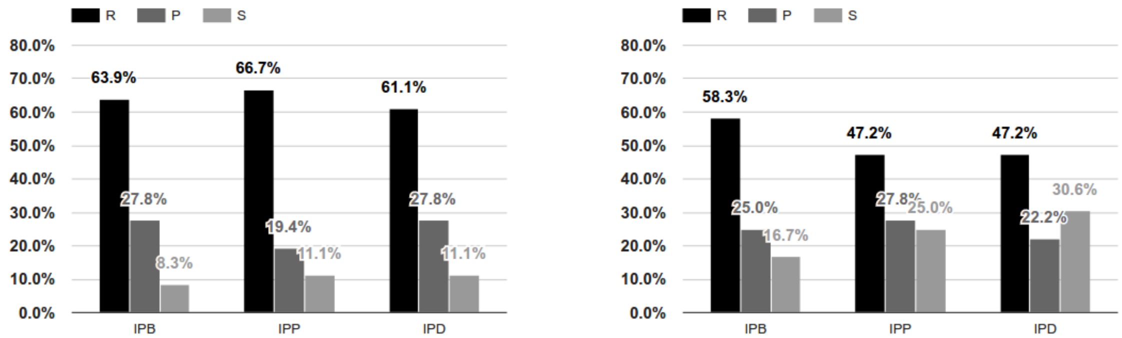
The *Mobile* columns of Table 4, similar to *Desktop*, show inconsistencies in the measures for search behavior. Participants saved documents in lower positions when relevant search results were placed lower (X^2 = 25.08, p < .001). We also observed that search performance improved by 24% from 42% to 52% as relevant search results were placed in higher positions from IPB to IPD conditions.

Environments & Search Behavior: Next, we examine the influence of different environments (*Desktop* and *Mobile*) on search behavior. Examining Table 3, the differences between the environments for search behavior measures can be determined. The p-values indicated on the tables are for the same conditions across

Table 4: Search behavior measures (M, SD) by Information Scent Pattern (ISP).

ISP	**Desktop**			**Mobile**		
Measures	**IPB**	**IPP**	**IPD**	**IPB**	**IPP**	**IPD**
TimeTotal$^{\#}$	268.22 (165.86)	**270.80 (162.70)**	259.22 (115.12)	211.91 (106.67)	211.86 (115.03)	**223.49 (162.83)**
Time$^{\#}$	72.67 (60.90)	62.57 (33.58)	**78.47 (49.12)**	**69.66 (50.05)**	67.49 (45.81)	67.54 (62.60)
NumQuery	**2.56 (2.10)***	2.51 (1.63)*	2.28 (1.83)	1.81 (.92)*	1.69 (.89)*	**1.86 (1.20)**
NumPage	**.56 (.50)**	.46 (.51)	.50 (.51)	**.44 (.50)**	.39 (.49)	.36 (.49)
NumClick	.50 (.94)	.60 (.88)	**.92 (1.32)**	.33 (.76)	.53 (.94)	**.58 (1.00)**
DRC	**3.53 (6.00)******	2.69 (4.96)	3.31 (5.42)**	1.86 (3.85)****	**2.56 (4.61)**	2.19 (4.57)**
DRV	**15.06 (5.29)****	14.17 (5.47)	14.75 (5.14)**	**12.64 (5.27)****	12.42 (5.59)	12.11 (5.46)**
ROA[1]	22%	17%	**11%**	19%	**8%**	**8%**
DRS	**5.06 (2.88)**	4.89 (3.62)	3.17 (2.29)	**5.50 (2.93)**	4.83 (3.30)	3.08 (1.66)
STotal	1.89 (1.47)	2.03 (1.42)	**2.44 (1.95)**	1.86 (1.31)	2.00 (1.29)	**2.19 (1.26)**
SRele	1.81 (1.43)	1.97 (1.42)	**2.19 (1.41)**	1.69 (1.33)	1.97 (1.23)	**2.08 (1.20)**
SRele%	45%	49%	**55%**	42%	49%	**52%**

Mean and (standard deviation) values are shown. Significant differences are indicated for same ISP conditions across different environments.
[1] Lowest value for ROA were bold for higher user interaction.
\# - indicates Student's t-test, otherwise Chi-squared test. Note: $*p < .05$, $^{*}_{*}p < .01$, $^{*}_{**}p < .001$, $^{**}_{**}p < .0001$.



(a) Desktop (b) Mobile

Figure 5: Distribution of Search Behavior by QueryAction: Reformulation (R), Pagination (P) and Stopping (S) for the first query controlled by ISP conditions on both desktop and mobile.

Table 5: Results for ISL/ISP conditions (X^2 significance) comparing search behavior across ISL/ISP conditions within the same environment (desktop or mobile).

	ISL		ISP	
Measure	**Desktop**	**Mobile**	**Desktop**	**Mobile**
TimeTotal	26.65****	16.41***	5.55	14.01***
Time	58.38****	113.19****	82.34****	1.08
NumQuery	0.43	0.99	0.64	0.29
NumPage	1.41	0.63	0.44	0.33
NumClick	3.85	7.00*	5.25	2.58
DRC	3.49	14.37***	5.23	3.94
DRV	3.77	1.21	2.10	0.41
ROA	2.60	4.00	1.37	2.46
DRS	65.11****	37.40****	17.12***	25.08****
STotal	39.20****	33.67****	3.07	1
SRele	37.62****	44.45****	1.47	1.51

$*p < .05$, $**p < .01$, $***p < .001$, $****p < .0001$

the two environments. For example, NumQuery for ILL on desktop and mobile is $p < .05$. Document click-throughs (DRC) are significantly higher on desktop compared to mobile across all the ISL conditions, when ISL is low ($X^2 = 18.75, p < .0001$), medium ($X^2 = 23.31, p < .0001$) and high ($X^2 = 4.52, p < .05$). More snippets are viewed (DRV) on desktop compared to mobile across all the ISL conditions, when ISL is low ($X^2 = 4.01, p < .05$), medium ($X^2 = 10.04, p < .01$) and high ($X^2 = 7.11, p < .01$). Considering lowest position of documents saved (DRS), participants saved significantly deeper on mobile compared to desktop for low ISL ($X^2 = 14.63, p < .001$). Some moderately significant differences in search behavior measures were observed between desktop and mobile when SERPs were manipulated under ISL conditions. Generally, in terms of differences between desktop and mobile, we recorded 5 notable differences: (1) query submissions numbered higher on desktop compared to mobile, significantly higher under ILL and ILH conditions. (2) Desktop participants significantly viewed more and (3) clicked on lower documents. (4) Mobile participants saved more results (significantly when there were fewer relevant results) from the first queries and (5) more accurately throughout all ISL conditions, compared to the desktop participants.

Differences in search behavior measures between desktop and mobile environments, under the influence of ISP conditions, are reported in Table 4. Desktop query submissions are significantly higher in number when relevant search results are lower in positions (IPB) ($X^2 = 4.64, p < .05$) or distributed throughout the SERPs (IPP) ($X^2 = 4.89, p < .05$). Document click-throughs (DRC) are also lower in position under IPB ($X^2 = 18.56, p < .0001$) and IPD ($X^2 = 8.08, p < .01$) conditions on desktop. More search results snippets are viewed on desktop under IPB ($X^2 = 7.59, p < .0001$) and IPD ($X^2 = 8.08, p < .01$) conditions. In general, search behavior measures between desktop and mobile, under Information Scent Pattern conditions, were not able to show any differences consistently.

5 DISCUSSION

We investigated the extent to which Information Foraging Theory could be used to explain changes in search behavior measures in different search environments. For both desktop and mobile, the findings suggest that ISL was a better predictor of search behavior than ISP. Allowing items to be saved as relevant directly on the SERP, without requiring click-throughs, we made different observations from previous work [30].

RQ1: ISL and Search Behavior: In RQ1, we sought to understand "to what extent can desktop and mobile search behavior be explained by Information Scent Level (ISL)". We posited that if ISL influenced search behavior, then the measures should increase correspondingly, apart from NumQuery and ROA which should be reducing because having enough relevant information should mitigate additional reformulations.

On desktop, apart from NumQuery, DRC and TimeTotal, there was a consistent increase in search behavior measures, when ISL increased from ILL to ILH. Documents saved (DRS, STotal and SRele) also significantly increased with ISL. We conclude that ISL was influential in desktop search behavior.

On mobile, the changes in user behavior measures were mixed. Increasing ISL did not increase search behavior measures consistently. Measurements for saved documents (DRS, STotal and SRele) and click depth (DRC) were significantly increased as ISL changed, but document click-throughs (NumClick) decreased from ILM to ILH. As documents saved (DRS, STotal and SRele) are the only measures that reflect ISL conditions, we concluded that ISL was only partially influential on mobile search behavior.

RQ2: ISP and Search Behavior: RQ2 sought to address "to what extent can desktop and mobile search behavior be explained by Information Scent Pattern (ISP)". We refer to ISP conditions by their ARS values in this subsection, as scent centrality is useful to explain how search behavior changes over the different conditions (see Table 1). If ISP influenced search behavior, then as ARS increased from 5.5 to 2.5, we would expect changes in search behavior that reflect user interactions. In alignment with previous work [12], we would expect document click-throughs (NumClick) and/or number of documents saved (STotal and SRele) to increase as the position of relevant documents were moved into higher positions. Rate of abandonment (ROA) and depth of document saved (DRS) were also be expected to decrease as it became easier to find relevant information [11]. We would also expect position-based measures (DRC, DRV and DRS) to be lower. Similarly, we would expect participants to expend less effort to reformulate, resulting in fewer query submissions (NumQuery).

On desktop, the identified search behavior measures mostly aligned with our hypothesized changes. As ARS improved from 5.5 to 2.5, NumClick, STotal and SRele increased while NumQuery and ROA decreased as expected. Overall, ISP was moderately successful to explain search behavior on desktop.

On mobile, only some search behavior measures agreed with our initial hypothesis. As ARS improved, document click-throughs (NumClick) and saved (STotal and SRele) increased likewise. Position-based measures, such as DRS and DRV decreased likewise. While not significant, changes in NumQuery and DRC were unexpected. DRC should be highest when ARS is 5.5, however, it was the lowest. This indicated that there was a higher probability that users opted not to click on anything, which resulted in the lowest value. When given a choice not to view documents, participants would select documents based on snippets alone. This observation was mentioned by Li et al. [21], discussing the significantly higher abandonment rate on mobile. This anomalous behavior on mobile will be discussed later.

RQ3: Environments and Search Behavior: RQ3 sought to address "how does search behavior differ as a result of different environments". If environments influenced search behavior, then search behavior measured across environments under identical ISL/ISP conditions, would be different. Results from Table 2 illustrate that measures were generally different between mobile and desktop.

Search strategies can be classified into two categories: depth-focused and reformulation-focused where depth and reformulation are inversely related [2]. While both NumQuery and DRV are higher on desktop than on mobile, the inverse is true for DRS (see Table 2). Mobile participants were more likely to find and save documents from the initial queries than desktop participants and avoid additional reformulations when possible. This behavior is also illustrated in Figures 4 and 5 where the Reformulation QueryAction was consistently lower across all ISL/ISP conditions on mobile than desktop. Search cost on mobile was thought to be higher because typing was harder.

The lower ROA on mobile than on desktop is consistent with previous work [21] as participants expend more effort to find relevant results within the first queries. Our user studies showed that under the ILH condition, mobile (but not the desktop) participants clicked on significantly fewer documents. Such a difference might suggest that information consumption satiety thresholds differ across the two environments.

We also observed that ranking affects search accuracy and confirmed *Guan and Cutrell*'s [11] findings. Search accuracy (SRele) increased for SERPs with more and higher ranked relevant results. The differences in search behavior on desktop and mobile were dependent on the type of Information Scent conditions. For tasks with an increasing number of relevant search results, mobile users had better search accuracy than desktop participants. Conversely, desktop users had better search accuracy than mobile participants for tasks with a distributed number of relevant search results. Overall, we found that different environments could affect changes in search behavior. The observation that NumClick was lower between ILM and ILH conditions only in the mobile environment may suggest a lower information need threshold, with the participants' information diet being restricted by the environment.

While visible search results above the fold influenced search behavior, having more relevant information below the fold should not make search behavior substantially different. Unlike Desktop, search measurements were not indicative that ILH had the highest information scent on mobile. Both NumPage and ROA were identical between the ILM and ILH conditions on mobile. ROA was also recorded as the same between the IPP and IPD conditions. A higher fold on mobile suggests a much diminished gain for including relevant information below the fold. Our study shows that user behavior on mobile is indeed different from desktop, similar to *Lagun et al. [20]*'s findings. However, the gap between mobile and desktop search is closing. It will be interesting to investigate how mobile search behavior continue to evolve.

Limitations. We acknowledge that an artificial time constraint of fourty-five minutes may potentially reduce the total number of documents examined [24]. However, the timing was kept consistent across both experiments. Search results are limited to twenty retrieved document, due to the low likelihood that users going beyond the first SERPs [30]. Participants spent less than five minutes per task on average and paginated very little in our experiments. We sought thirty-six participants for each user study and noted that other user studies had smaller [14, 18, 20] or similar [10, 30] numbers of participants. As SERPs were prepared beforehand, we recognized that participants might encounter SERPs that were not targeted to their initial query terms but there were strong merits to keep SERPs consistent to users. Participants were interviewed during the exit questionnaires, and apart from one user, no concerns were raised regarding number of search results, total time given, and SERP manipulations. From this evidence, we conclude the experimental manipulations were not noticeable in general.

6 CONCLUSION AND FUTURE WORK

This research investigated how ISL and ISP can be used to measure differences in web search behavior in mobile and desktop environments. We found that desktop participants behaved in similar ways to those observed in past work [30] but not for mobile participants.

By allowing participants to save answer items directly on the SERP, without having to examine documents, we observed that document click-throughs were not an indicator for the strength of information scent level. This is relevant for mobile, because of previously observed higher good abandonment rate [21]. In general, participants in both environments tended to abandon SERPs when the number of relevant search results were fewer, or if found after non-relevant search results. Users were also more likely to click documents in lower positions when more relevant search results were present on the SERPs.

While participants consistently preferred SERPs with a higher number of relevant search results on desktop, this preference was not apparent on mobile. We conjectured that the higher fold on mobile impaired their initial impression of differences in overall page quality, since they were only able to see the first few items, but more research is required to fully understand this effect. Desktop participants submitted more queries and saved fewer documents in lower positions than their mobile counterparts. Differences in information scent and environments have been observed to change search behavior. The significant inverse relationship between NumQuery and DRV in different environments suggested that whether the search was carried out on desktop or mobile influenced their search strategies. These differences in preferences may also contribute to how information is consumed in different environments.

In conclusion, we conducted two comparative user studies using IFT and found differences between two environments. Increasing ISL generally increased search interactions under desktop ISL conditions. However, NumClick dropped when ISL was above ILM under mobile ISL conditions. A possible lower information need threshold in mobile environment has been suggested. Similar to previous work, the results under ISP conditions for both environments were mixed. In terms of search strategies, an inverse relationship between NumQuery and DRS was observed. This suggests that search strategies might change, contingent on the environment. Our findings have implications for the design of search systems and suggest several areas for future investigation to improve search accuracy on mobile: 1) presenting more search results with shorter snippets above the fold, 2) techniques to make query reformulation easier, and 3) considering approaches such as tiered snippets to present additional information within limited space.

7 ACKNOWLEDGMENTS

This project is funded by ARC Discovery Grant, ref: DP140102655 and an APA scholarship. Travel funding is also provided by ACM SIGIR for the lead author to attend the conference. We thank Diane Kelly and Wan-Ching Wu for providing clarification as well as Bruce Croft, Doug Oard and the anonymous reviewers for their valuable feedback.

REFERENCES

[1] Anne Aula, Päivi Majaranta, and Kari-Jouko Räihä. 2005. Eye-Tracking Reveals the Personal Styles for Search Result Evaluation. In *Proceedings of INTERACT.* 1058–1061.
[2] Leif Azzopardi, Diane Kelly, and Kathy Brennan. 2013. How Query Cost Affects Search Behavior. In *Proceedings of SIGIR.* 23–32.
[3] Andrei Broder. 2002. A Taxonomy of Web Search. In *SIGIR Forum*, Vol. 36. 3–10.
[4] Stuart K. Card, Peter Pirolli, Mija Van Der Wege, Julie B. Morrison, Robert W. Reeder, Pamela K. Schraedley, and Jenea Boshart. 2001. Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability. In *Proceedings of SIGCHI.* 498–505.
[5] Ed H. Chi, Peter Pirolli, Kim Chen, and James Pitkow. 2001. Using Information Scent to Model User Information Needs and Actions and the Web. In *Proceedings of SIGCHI.* 490–497.
[6] Ed H. Chi, Peter Pirolli, and James Pitkow. 2000. The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site. In *Proceedings of SIGCHI.* 161–168.
[7] Karen Church and Barry Smyth. 2009. Understanding the Intent behind Mobile Information Needs. In *Proceedings of IUI.* 247–256.
[8] Edward Cutrell and Zhiwei Guan. 2007. What Are You Looking For?: An Eye-Tracking Study of Information Usage in Web Search. In *Proceedings of SIGCHI.* 407–416.
[9] Anindya Ghose, Avi Goldfarb, and Sang Pil Han. 2012. How Is the Mobile Internet Different? Search Costs and Local Activities. *Information Systems Research* 24, 3 (2012), 613–631.
[10] Laura A. Granka, Thorsten Joachims, and Geri Gay. 2004. Eye-Tracking Analysis of User Behavior in WWW Search. In *Proceedings of SIGIR.* 478–479.
[11] Zhiwei Guan and Edward Cutrell. 2007. An Eye Tracking Study of the Effect of Target Rank on Web Search. In *Proceedings of SIGCHI.* 417–420.
[12] Ahmed Hassan, Xiaolin Shi, Nick Craswell, and Bill Ramsey. 2013. Beyond Clicks: Query Reformulation as a Predictor of Search Satisfaction. In *Proceedings of CIKM.* 2019–2028.
[13] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. Accurately Interpreting Clickthrough Data as Implicit Feedback. In *Proceedings of SIGIR.* 154–161.
[14] Matt Jones, Gary Marsden, Norliza Mohd-Nasir, Kevin Boone, and George Buchanan. 1999. Improving Web Interaction on Small Displays. *Computer Networks* 31, 11 (1999), 1129–1137.
[15] Maryam Kamvar and Shumeet Baluja. 2006. A Large Scale Study of Wireless Search Behavior: Google Mobile Search. In *Proceedings of SIGCHI.* 701–709.
[16] Maryam Kamvar and Shumeet Baluja. 2007. Deciphering Trends in Mobile Search. *Computer* 40, 8 (2007), 58–62.
[17] Jaewon Kim, Paul Thomas, Ramesh Sankaranarayana, and Tom Gedeon. 2012. Comparing Scanning Behaviour in Web Search on Small and Large Screens. In *Proceedings of ADCS.* 25–30.
[18] Jaewon Kim, Paul Thomas, Ramesh Sankaranarayana, Tom Gedeon, and Hwan-Jin Yoon. 2015. Eye-Tracking Analysis of User Behavior and Performance in Web Search on Large and Small Screens. *JASIST* 66, 3 (2015), 526–544.
[19] Kerstin Klöckner, Nadine Wirschum, and Anthony Jameson. 2004. Depth-and Breadth-First Processing of Search Result Lists. In *Proccedings of CHI.* 1539–1539.
[20] Dmitry Lagun, Chih-Hung Hsieh, Dale Webster, and Vidhya Navalpakkam. 2014. Towards Better Measurement of Attention and Satisfaction in Mobile Search. In *Proceedings of SIGIR.* 113–122.
[21] Jane Li, Scott Huffman, and Akihito Tokuda. 2009. Good Abandonment in Mobile and PC Internet Search. In *Proceedings of SIGIR.* 43–50.
[22] David Maxwell, Leif Azzopardi, Kalervo Järvelin, and Heikki Keskustalo. 2015. Searching and Stopping: An Analysis of Stopping Rules and Strategies. In *Proceedings of CIKM.* 313–322.
[23] Peter Pirolli and Stuart Card. 1999. Information Foraging. *Psychological Review* 106, 4 (1999), 643–675.
[24] Chandra Prabha, Lynn Silipigni Connaway, Lawrence Olszewski, and Lillie R. Jenkins. 2007. What Is Enough? Satisficing Information Needs. *Documentation* 63, 1 (2007), 74–89.
[25] Yongli Ren, Martin Tomko, Kevin Ong, and Mark Sanderson. 2014. How People Use the Web in Large Indoor Spaces. In *Proceedings of CIKM.* 1879–1882.
[26] Soo Young Rieh and others. 2006. Analysis of Multiple Query Reformulations on the Web: The Interactive Information Retrieval Context. *IP&M* 42, 3 (2006), 751–768.
[27] Yang Song, Hao Ma, Hongning Wang, and Kuansan Wang. 2013. Exploring and Exploiting User Search Behavior on Mobile and Tablet Devices to Improve Search Relevance. In *Proceedings of WWW.* 1201–1212.
[28] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. 2004. The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of SIGCHI.* 415–422.
[29] Ryen White and Steven M. Drucker. 2007. Investigating Behavioral Variability in Web Search. In *Proceedings of WWW.* 21–30.
[30] Wan-Ching Wu, Diane Kelly, and Avneesh Sud. 2014. Using Information Scent and Need for Cognition to Understand Online Search Behavior. In *Proceedings of SIGIR.* 557–566.