Comparing scanning behaviour in web search on small and large screens

Jaewon Kim† Paul Thomas§ Ramesh Sankaranarayana† Tom Gedeon†

†Research School of Computer Science
The Australian National University
{jaewon.kim, ramesh.sankaranarayana, tom.gedeon}@anu.edu.au

§CSIRO ICT Centre
Canberra, Australia
Paul.thomas@csiro.au

ABSTRACT
Although web search on mobile devices is common, little is known about how users read search result lists on a small screen. We used eye tracking to compare users' scanning behaviour of web search engine result pages on a small screen (hand-held devices) and a large screen (desktops or laptops). The objective was to determine whether search result pages should be designed differently for mobile devices. To compare scanning behaviour, we considered only the fixation time and scanning strategy using our new method called 'Trackback'. The results showed that on a small screen, users spend relatively more time to conduct a search than they do on a large screen, despite tending to look less far ahead beyond the link that they eventually select. They also show a stronger tendency to seek information within the top three results on a small screen than on a large screen. The reason for this tendency may be difficulties in reading and the relative location of page folds. The results clearly indicated that scanning behaviour during web search on a small screen is different from that on a large screen. Thus, research efforts should be invested in improving the presentation of search engine result pages on small screens, taking scanning behaviour into account. This will help provide a better search experience in terms of search time, accuracy of finding correct links, and user satisfaction.

Keywords
Scanning behaviour, small screen, Trackback

1. INTRODUCTION
Accessing the web on mobile devices is becoming increasingly popular.[1] In this context, the following question is important: on small devices, what is users' scanning behaviour when they search for information on the web? In this study, scanning behaviour is defined as the users' actions on each element of the search engine result pages, such as seeking strategies, fixation, scanpath, click, or scroll. This question is of interest because understanding scanning behaviour is invaluable for improved interface design or for obtaining more targeted metrics for evaluating the retrieval performance [3, 5, 8, 12]. If there is any difference in users' scanning behaviour on small and large screens, then we should consider designing the presentation of results differently for each sized devices.

One method for understanding scanning behaviour is analysing transaction log files that have information about click-through, queries, and scrolling interactions between users and search engines [4, 9, 13, 17]. Another approach uses diary studies to investigate the use of web search engine with individual interviews [18]. Beyond these earlier studies, eye tracking seems to facilitate our understanding of users' attention, because their gaze can show when they are paying attention to elements of web search engines, moment by moment [3, 5, 7, 8].

Much research has been conducted on users' scanning behaviour by using eye tracking to determine where and how people look at web search results. Several studies have classified users' scanning behaviour according to gaze patterns. Klöckner et al. [11] found that 52–65% of participants used what they call a 'depth-first' strategy (the subjects scanned only the links above the selected link), 11–15% used a 'breadth-first' strategy (the subjects looked through all the links before making a decision and selecting a link), while the remaining 20–37% showed a 'mixed' strategy (looking ahead a few results past the selected link). Aula et al. [1] defined two kinds of evaluation patterns. They suggested that 54% of subjects who scanned less than half of the visible results were 'economic' evaluators, and that the others had an 'exhaustive' evaluation style. Dumais et al. [7] extended the classification and defined three clusters—Economic-Results, Economic-Ads, and Exhaustive—to identify users' scanning patterns when viewing the results of major commercial search engines that include additional links such as sponsored links or advertisements. According to their results, the Economic-Results and the Economic-Ads groups tended to spend more time on the first three results than did the Exhaustive users (68%, 61%, and 53%). In

ADCS '12, December 5–6, 2012, Dunedin, New Zealand.

[1] A report from 2010 indicates that even though desktop PCs or laptops are used, more than 40% of Australians mainly use mobile phones to access the web.

Source: "Australian mobile internet usage doubles", http://thenextweb.com/au/2010/09/23/australian-mobileinternet-usage-doubles, (Retrieved May. 06, 2012).

addition, the total fixation time of each group showed that the Exhaustive participants reviewed the results most slowly.

Numerous researchers have investigated scanning of behaviour features to study how users normally scan the elements of search results. Several research studies have examined eye fixations related to scrolling, and scanning behaviour above and below the selection [8, 10]. Their results suggest that subjects rarely scan below the selected link except when the link is at the page fold (when users often scan further), and that the scanning direction is from top to bottom. Buscher et al. [3], who used behavioural log data in a commercial web search engine instead of eye tracking, found that users who spend shorter time on search result pages tend to inspect just a few results, scroll less, and use fast mouse movement. Lorigo et al. [12] described two types of tasks, informational and navigational, that may impact task completion time. They found that users tend to spend more search time on informational than on navigational tasks.

Given that we have to display search results differently on small screens, if users' scanning behaviour is different from that on large-sized screens, there may be other implications for the display of search results. Only a few studies of eye tracking on small screens have been performed. Drewes et al. [6] investigated gaze interaction for controlling applications on a handheld device using the dwell-time method and gaze gestures. Further, Nagamatsu et al. [14] investigated a remote gaze tracker for small devices with stereo gaze tracking. Recently, text interaction and reading performed on an actual mobile touch screen device was analysed by Biedert et al. [2]. However, no study has compared users' scanning behaviour during web search on small and large screens. Therefore, it is not clear how users read typical displays of search results on mobile devices, and whether these displays can be improved.

In this study, we conducted an eye-tracking study with 32 participants who completed 20 tasks on large and small screens and we focused on the relation between scanning behaviour and the screen size by resizing the web browser. We measured fixation time to investigate users' search performance and attention, adopted one classification method from previous work, and defined a method we call 'Trackback' to see how far subjects look ahead prior to making the first selection.

2. EXPERIMENT

With an eye-tracking instrument that provides users' scanning behaviour of web search result pages, we recorded gaze data for each of the 32 participants who completed search tasks on a large as well as a small display screen.

2.1 Tasks

Each participant completed 20 search tasks. Inspired by Lorigo et al. [12], we used 10 tasks of each of the two kinds of tasks (informational and navigational) to investigate any influence of the task type. Each task showed a task description and a predefined query obtained from Dumais et al. [7] modified to optimize local participants' understanding (see Table 1), and search results were retrieved from the Google mobile search engine, from which we removed images, maps, or related links so that all tasks showed the same kinds of elements. All results pages were cached as local files in the system and relevant pages were shown when subjects chose the links. The tasks were very simple, needing only 1–2 min to complete. According to the search results from the Google search engine, 18 of the predefined result pages contained a relevant solution within the top 3 results, with the other two including a relevant result in ranks 4–6.

Table 1: Examples of task descriptions and queries. Nav denotes navigational task and Info denotes informational task.

Task Description	Initial Task Query	Task Type
Find the official homepage of the Canberra casino and hotel in Canberra.	Canberra Casino	Nav
Go to the homepage of the Canberra Cavalry baseball team.	Canberra cavalry baseball	Nav
What is the standard length of a cue used for playing billiards?	billiard cue size	Info
How many spikes are in the crown of the Statue of Liberty?	statue of liberty crown spikes	Info

2.2 Design

The participants were divided into four groups of eight, and the tasks were arranged in two sets: set 1 consisted of informational tasks 1 to 5 and navigational tasks 1 to 5, and set 2 consisted of the remaining tasks (i.e., set 1 consisted of I1N1I2N2I3N3I4N4I5 N5 and set 2 of I6N6I7N7I8N8I9N9I10N10, where 'I' denotes an informational and 'N' a navigational task). Each subject performed both task sets, one on a large screen and one on a small screen, and both the set order and screen order were counterbalanced across subjects. In other words, subject 1 performed task set 1 (TS1) on the large screen and then task set 2 (TS2) on the small screen, followed by subject 2 performing TS2 on the large screen and then TS1 on the small screen, and so on (see Table 2). Therefore, each task was distributed 32 times (16 times on each screen size) over the participants. In other words, the eight subjects in each group performed the tasks in exactly the same order, and faced the two screen sizes in the same sequence. Finally, we assigned 10 areas of interest (AOIs) on each search result page to investigate users' gaze and fixation. Each AOI corresponded to a search result, i.e., a clickable link along with its snippet text and a URL.

Table 2: Examples of design for each group. L denotes a large screen and S denotes a small screen.

	Task set, order and screen size
Group 1	TS1 on L, and then TS2 on S
Group 2	TS2 on L, and then TS1 on S
Group 3	TS1 on S, and then TS2 on L
Group 4	TS2 on S, and then TS1 on L

2.3 Procedure

First, all participants listened to an introduction to the experiment, and then practiced two sample tasks on each size screen until they were familiar with the system. Their head was then fixed on a chinrest to ensure higher eye gaze detection accuracy and the eye tracker was calibrated using 5-point calibration. Next, we preceded the first task description with an initial query and then showed the result page; this procedure was repeated until task number 20 according to an automated schedule. A time notice of 3 min was given after starting each task, after which the subjects were free to either spend more time to find the answer or move on to the next task. Typing a query was not allowed to prevent the subjects from looking at the keyboard. However, they could continue to the next page of results or follow links from the list of results. The participants were allowed to ask for task description if they did not understand the task sufficiently. At the end of the experiment, the subjects were asked to complete a questionnaire about their web search experience (the questionnaire results are still being analysed, and are not discussed in this paper). The experimental run time was approximately 30 min for each participant.

2.4 Participants

35 subjects (19 male) between 18 and 50 years, from various disciplines and recruited on campus at a local university, participated in the eye-tracking study. All subjects were experienced in web searching and were very familiar with the Google search engine. We excluded the results from three participants for technical reasons (e.g., stability problems with eye tracking).

2.5 Experimental setup

All search results were obtained from the Google mobile search engine and displayed in Internet Explorer 8. Creating new tabs or new windows was prohibited. Eye gaze was recorded by Facelab 5 with a desk mounted 17" LCD monitor and with a chinrest, and data analyses were performed using Eyeworks software.[2]

The large screen ran at 1280 × 1024 pixels as the default resolution of the monitor. Because of users' movements and the obstruction of the camera by the hand used for touching a screen, it is difficult to record the point at which a user is looking [14]. To simulate the small screen of a mobile device, we used the same monitor but with a browser limited to a 320 × 480 pixel window. To compare the effects of different screen sizes, we adopted the same font size on both types of browsers. With these settings, in the case of the large screen there was no fold (see Figure 1) and the ten search results were clearly visible without needing any

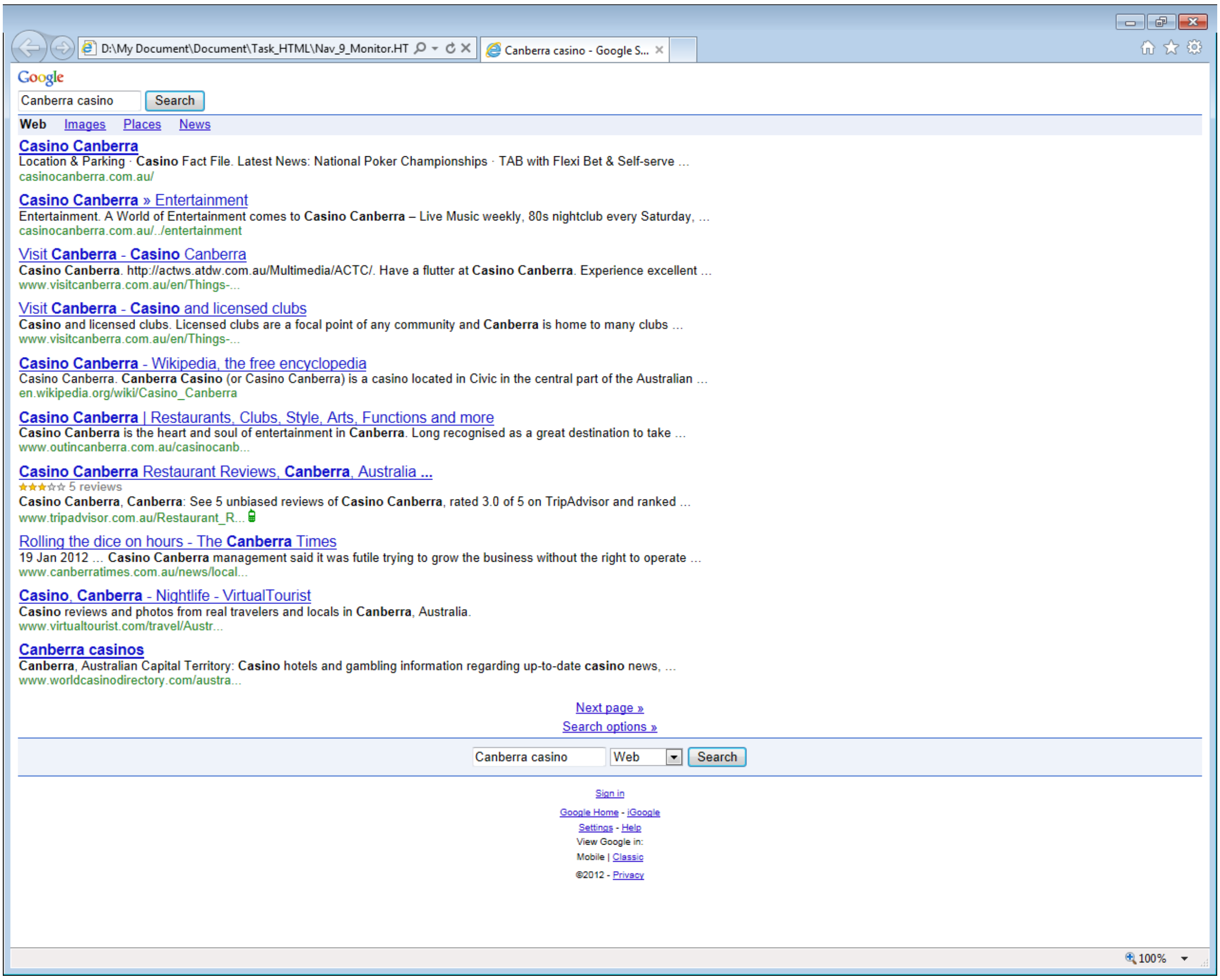

Figure 1. Task on the large screen.

user control of the browser, whereas the page fold was normally between positions 3 and 4 of the search results on the small screen (see Figure 2).

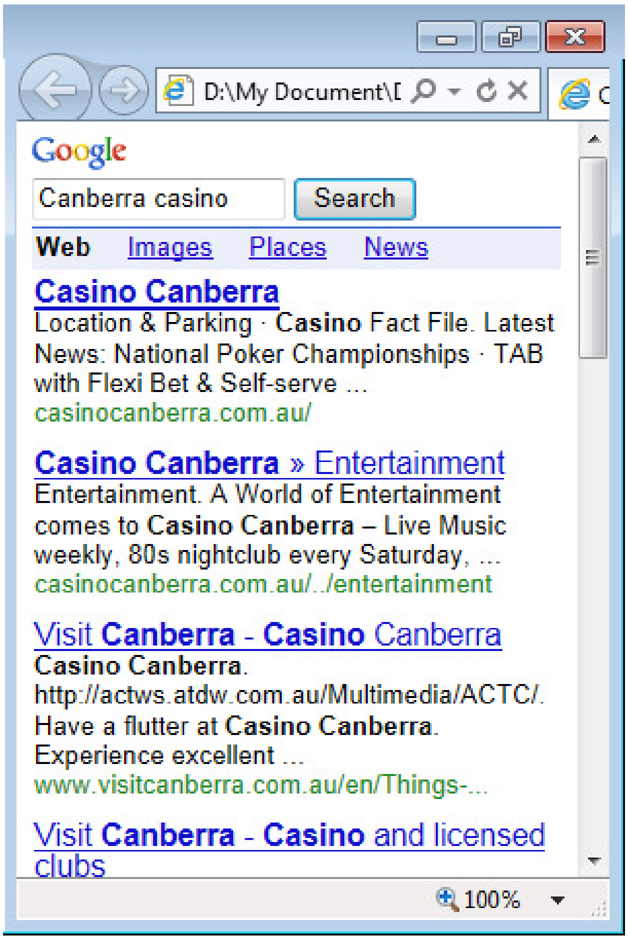

Figure 2. Task on the small screen.

3. RESULTS

Our data consist of gaze data from 640 queries (320 queries on the large screen and the same on the small screen, 160 queries of each informational and navigational task on each sized screen.) We adopted all of the 32 users' data for the results. In this paper, we focus only on fixation time, scanning strategies up to the first click, and Trackback, which we defined to describe how much users scan beyond the selected links. The other data, such as task completion time, saccades, scanpaths, and the questionnaire, are still being analysed. To study users' eye gaze in detail with the relatively small font sizes of the web search result pages on a mobile search engine, the fixations were recorded if a gaze lasted at least 75 ms and if the gaze locations were close to each other (within a radius of five pixels) using the built-in algorithms of Eyeworks software.

[2] The website: http://www.seeingmachines.com/product/facelab/

3.1 Fixation time

Eye fixations are useful for comparing users' scanning behaviours, because they indicate the point at which the user is looking and the fixation time represents the user's interest in each AOI or the difficulty of tasks [15, 16]. A comparison among the total fixation durations on each AOI may provide information about the usability of the interface as well as the users' efforts and search performance. The results showed slight differences between screen sizes and task types: the total fixation time up to the first click on the small screen was about 15% longer than on the large screen: 1337 s versus 1166 s, although this is not statistically significant with ANOVA $F = 3.36$, $p = 0.07$. The average fixation time for the informational tasks was slightly longer than for the navigational tasks on both sizes of screen: 3.91 s and 3.38 s on the large screen, and 4.56 s and 3.8 s on the small screen. These differences are also not statistically significant. This difference was particularly pronounced considering the top ranks: the total time spent on AOIs 1 to 3 for navigational tasks on the small screen was about 20% (100 s) shorter than it was for the informational tasks. The results according to screen sizes and task effects imply that users can easily find the link they need on a large screen in navigational searches.

Figure 3 shows the results when normalizing for total gaze time. After the first link, the AOI-normalized percentages on both screen sizes decrease sharply. However, the proportion of time spent on the first AOI on the small screen was much higher than that on the large screen (48% versus 39%) and fixations on the periphery, e.g., a query box, category tabs, or blanks between AOIs, on the large screen showed about 7% more subject attention. Even if we do not consider the proportion on the periphery, the proportion of AOIs 1 and 2 was about 4% higher on the small screen, whereas the proportions of all the other AOIs were higher on the large screen. Participants tended to spend more time on results ranked 1 to 3 when using the small screen than when using the large screen (76% on the small screen versus 67% on the large screen). This result indicates that the links ranked more than three on the small screen received very little users' attention, even though they spent a longer time overall on the small-sized screen.

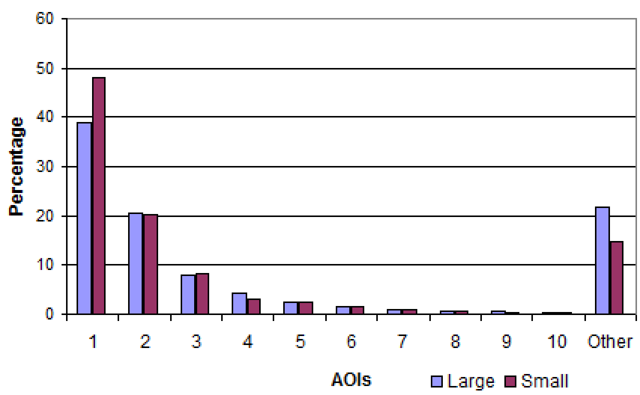

Figure 3. AOI-normalized time spent viewing each AOI [%]

3.2 Scanning strategies

We examined scanning strategies for the initial pages of search results. Even though the classification of Aula et al. [1], i.e., economic and exhaustive evaluators, seemed to be well defined and adequate for providing more invaluable comparisons, we decided to adopt the strategies of Klöckner et al. [11], because Aula et al.'s strategy did not seem suitable for the few visible links on a small screen. If there are a few search result links, such as three or four, an economic evaluator would be defined as someone who scans only one or two of the search results.

The depth- and breadth-first strategies of Klöckner et al. [11] are useful abstractions of users' decision patterns: 'depth-first' users follow a promising link immediately; 'breadth-first' users read all their options exhaustively before clicking; and 'mixed' users read ahead but to a smaller extent. We adopted this distinction when analysing our data. Table 3 shows the total count and proportion of participants' scanning behaviours differentiated by the three kinds of strategies recognised by this approach. The table shows that subjects tended to use the depth-first strategy on the large screen slightly less than on the small screen (116 versus 131, i.e., 36% versus 41%). Instead, on the large screen they used the breadth-first strategy twice as much as on the small screen. The distributions of strategies are significantly different across screen sizes ($\chi^2 = 11.89$, df = 2, $p < 0.01$), but the count and proportion of the mixed strategy is almost the same on both screen sizes.

Table 3. Choice of scanning strategy on both screen sizes

	Large			Small		
	Depth	Mixed	Breadth	Depth	Mixed	Breadth
Total	116	184	20	131	179	10
%	36	58	6	41	56	3

This result seems to imply that there is not a big difference in users' strategy between screen sizes. However, the majority of cases are 'mixed', with some degree of reading ahead, and therefore, this classification is perhaps hiding some differences.

3.3 Trackback

To examine the behaviour as the 'mixed' strategy in greater detail, we define 'Trackback' as the difference in ranks between the selected link and the farthest link observed. For example, if a subject looked as far as AOI 7 and then clicked AOI 3, the Trackback value is 4. We were able to consider only the farthest link because all our users scanned from top to bottom as in the previous study [8, 10]. With Trackback, we can scrutinize differences within the mixed strategy; the higher the Trackback, the greater is the extent to which links are observed. This method has some similarity to the scanpath analysis method used in a previous study [7, 12, 15]. However, Trackback is unique in that it summarizes the amount of additional effort users make before selecting links.

Across all users, overall there is a very significant difference; the Trackback value on the large screen is about 54% higher than on the small screen (mean 1.95 ranks/user/task versus 1.27; $t = 3.78$, df = 601, $p < 0.001$).

To examine the change in the Trackback value from large to the small screens, we calculated each participant's difference in the Trackback value between the large and small screens: the difference in Trackback for each user equals the sum of Trackback values on the large screen minus the total Trackback value on the small screen. Figure 4 illustrates the difference for each participant. Points above the x-axis represent a higher Trackback value (more looking ahead) on the large screen and points below the x-axis represent higher Trackback on the small screen. 21 users have high Trackback on the large screen whereas only 11 have high Trackback on the small screen. Therefore, screen size certainly has an effect on Trackback (ANOVA $F = 13.01$, $p < 0.001$), thereby affecting scanning behaviour. Thus, the Trackback value on a large screen is normally higher than that on a small screen.

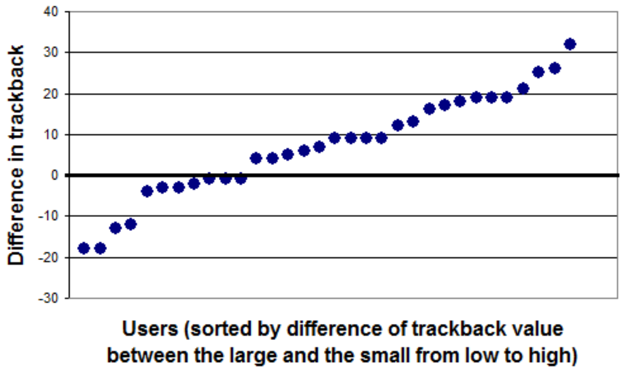

Figure 4. Distribution of difference in Trackback between both sizes of screen. Points above the x-axis represent higher Trackback on the large screen.

4. DISCUSSION AND FUTURE WORK

We have presented a study to investigate how users' scanning behaviour is different between a small and a large screen. This study showed differences in users' time spent and, by using the Trackback method, their scanning strategies on each screen size. First, from the AOI fixations in our data, we can see that users tended to take slightly longer time to decide the first selections on the small screen, although the difference is not significant. If this effect holds up under further investigation, it is probably because the small screen interface is less comfortable than the large one. In addition, the need to scroll to see more results affects the fixation time. Conversely, the difference between task types is clear: the informational tasks take more effort and time, as found in a previous study [12]. Moreover, the participants showed higher proportions of fixation time on the top three AOIs on the small screen than on the large screen, but there was no such effect associated with task type. This means that the proportions on each AOI are not influenced by task types, but by screen size. This can also be explained by the fact that the page fold is located around results 3 to 4 on the small screen, and users hardly use the scroll bar: 76% of fixations on the small screen were within the top three results, and only 9% beyond these.

Second, in the comparison between the small and large screens in terms of depth- and breadth-first strategies, users implemented more breadth-first strategy and less depth-first strategy on the large screen than on the small screen. This seems to be because of the scrolling required on the small screen, with fewer lists of results showing on the initial screen, i.e., ten results versus three or four results on the large screen versus the small screen, respectively. The classification is not entirely useful since so many participants use a "mixed" strategy and this hides real differences.

Lastly, the results of the Trackback method for observing the difference in users' choice of mixed-strategy between the large and small screen show that the average Trackback on the small screen is 1.27 per task, whereas it is 1.95 on the large screen. In addition, only about 13% of subjects showed a large negative Trackback, where the sum of Trackback values is less than –10. This indicates that subjects tend to look over more items on the large screen. We believe this means that on a large screen, user gather more information before selecting a result: they may be being more careful and checking their selection before committing to it.

There is definitely a difference in users' scanning behaviour on differently sized screens. Therefore, we should contemplate improving the presentation of web search engine result pages on small devices separately to provide users' with better search experience, even though several kinds of users' scanning behaviour have been studied on large-sized screens. The results may suggest that web interface designers or developers need to investigate the optimum presentation of search result pages for the small screen to facilitate less scanning as well as fast search time.

In further studies, first, we plan to analyse in detail the data of this experiment to reveal the difference in scanning behaviour on both screen sizes. Subsequently, since this experiment demonstrated that users tend to spend more time in spite of less scanning on a small screen, our next step will suggest an improved presentation design of web search engines on small devices by studying the relations between visible contents such as snippets, URL, or font sizes and users' scanning behaviour.

5. REFERENCES

[1] Aula, A., Majaranta, P. and Raiha, K-J. 2005. Eye-tracking reveals personal styles for search result evaluation. In *proceeding of the 2005 IFIP TC13 international conference on Human-Computer Interaction* (INTERACT'05), Maria Francesca Costabile and Fabio Paternò (Eds.). Springer-Verlag, Berlin, Heidelberg, 1058-1061.

[2] Biedert, R., Dengel, A., Buscher, B. and Vartan, A. 2012. Reading and estimating gaze on smart phones. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA '12), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 385-388.

[3] Buscher, G., Dumais, S. and Cutrell, E. 2010. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '10). ACM, New York, NY, USA, 42-49.

[4] Cohen, W., Shapire, R., and Singer, Y. 1999. Learning to order things. In *Journal of Artificial Intelligence Research,* 10, 243–270.

[5] Cutrell, E. and Guan, Z. 2007. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (CHI '07). ACM, New York, NY, USA, 407-416.

[6] Drewes, H., De Luca, A., and Schmidt, A. 2007. Eye-gaze interaction for mobile phones. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology* (Mobility '07). ACM, New York, NY, USA, 364-371.

[7] Dumais, S., Buscher, G., and Cutrell, E. 2010. Individual differences in gaze patterns for web search. In *Proceedings of the third symposium on Information interaction in context* (IIiX '10). ACM, New York, NY, USA, 185-194.

[8] 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* (SIGIR '04). ACM, New York, NY, USA, 478-479.

[9] Jansen, B. J. and Spink, A. 2006. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*, 42(1), 248–263.

[10] 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* (SIGIR '05). ACM, New York, NY, USA, 154-161.

[11] Klöckner, K., Wirschum, N. and Jameson, A. 2004. Depth- and breadth-first processing of search result lists. In *CHI '04 extended abstracts on Human factors in computing systems* (CHI EA '04). ACM, New York, NY, USA, 1539-1539.

[12] Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L. and Gay, G. 2006. The influence of task and gender on search evaluation and behavior using Google. *Information Processing and Management*, 42(4), 1123-1131.

[13] Mat-Hassan, M. and Levene, M. 2005. Associating search and navigation behavior through log analysis. In *Journal of the American Society for Information Science and Technology*, 56(9), 913-934.

[14] Nagamatsu, T., Yamamoto, M., and SATO, H. 2010. MobiGaze: development of a gaze interface for handheld mobile devices. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* (CHI EA '10). ACM, New York, NY, USA, 3349-3354.

[15] Pan, B., Hembrooke, H.A., Gay, G.K., Granka, L.A., Feusner, M.K., 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* (ETRA '04). ACM, New York, NY, USA, 147-154.

[16] Poole, A., & Ball, L. J. 2005. Eye tracking in human-computer interaction and usability research: Current status and future. In *Prospects, Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc.*

[17] Silverstein, C., Henzinger, M., Marais, H. and Moricz, M. 1998. Analysis of a very large AltaVista query log. SRC Technical note #1998-14. On-line at http://gatekeeper.dec.com/pub/DEC/SRC/technicalnotes/abstracts/src-tn-1998-014.html.

[18] Teevan, J., Alvarado, C., Ackerman, M.S., and Karger, D.R. 2004. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (CHI '04). ACM, New York, NY, USA, 415-422.