Sorting out Searching on Small Screen Devices

Matt Jones [1], George Buchanan[2] & Harold Thimbleby[3]

[1]Dept. of Computer Science, University of Waikato, New Zealand
always@acm.org
[2]Interaction Design Centre Middlesex University, London, UK.
george10@mdx.ac.uk
[3]UCLIC, University College London Gower St, London UK
H.Thimbleby@cs.ucl.ac.uk

Abstract. Small handheld devices – mobile phones, PDAs etc – are increasingly being used to access the Web. Search engines are the most used Web services and are an important user support. Recently, Google™ (and other search engine providers) have started to offer their services on the small screen. This paper presents a detailed evaluation of the how easy to use such services are in these new contexts. An experiment was carried out to compare users' abilities to complete realistic tourist orientated search tasks using a WAP, PDA-sized and conventional, desktop interface to the full Google™ index. With all three interfaces, when users succeed in completing a task, they do so quickly (within 2 to 3 minutes) and using few interactions with the search engine. When they fail, though, they fail badly. The paper examines the causes of failures in small screen searching and proposes guidelines for improving these interfaces.

1 Introduction

Research into usable, useful and effective approaches for users to search the Web is vital. Unless effective user-centered approaches are developed and applied, the promise of wide access to the information resources will be lost, with users left frustrated and overwhelmed [16]. There is much work on the social impact of small screen devices [e.g., 13], which more-or-less take the technology and the user interface design in particular as given; this paper, in contrast, reviews and discusses experiments on the usability impacts of software and structural aspects of user interface design on small screens for supporting web applications.

Recently, Google™, one of the most comprehensive Web search engines, has offered its services to WAP phone users and via PDA type handheld computers. We evaluated the usability of these services to explore the impact of screen size. We looked at the performance of users under realistic search task situations on the different platforms. We were interested in differing patterns in user behaviour with the three services studied.

On a typical desktop screen, the user has many different ways to interact, often with varied interaction styles (menus, direct manipulation, text, etc). The rich desktop environment is in contrast to the "impoverished" interfaces of mobile, handheld devices. While a range of interesting search result visualization and manipulation schemes have been proposed for large screen devices (e.g., [6]), these schemes, on the

whole, are not appropriate to handheld devices. Our work presents some first guidelines for those involved in designing approaches that are more appropriate for the small screen contexts.

2. Background and Review

2.1 Interfaces for Mobile Web Browsing

Most Web pages are designed for conventional large screen viewing. In earlier work we assessed the impact on user interaction of using such pages with the small display areas found on handheld computers. The study [11] suggested that users did not want to use the conventional page-to-page navigation as it was interactively very costly on the small screen. Rather, a much more direct, systematic approach requiring less scrolling was seen as appropriate.

WebTwig [12] was developed to demonstrate a direct approach to handheld browsing that takes account of the limited display. The tool presents an hierarchical outline view which users can manipulate as they attempt to identify useful areas of the Web site. User evaluations of our system suggest benefits of this approach. Recent work by others amplifies our findings [2].

2.2 Interfaces for Searching

Most search systems simply present the results of a user query as a (long) ranked list, often broken into pages, of matching documents. With such an interface users have to scroll and page through the often-overwhelming list, examining documents in detail as they proceed to make relevance judgements. Such approaches mean that even on conventional large displays search interfaces are not highly usable. As Shneiderman *et al* [16] put it, "…the result is confusion and frustration."

Hearst [9] has identified two key types of interaction search interfaces should support. First, users should be able to scan search results quickly getting a feel for the effectiveness of their query and the sorts of information available. Second, interfaces must also facilitate the flexible, dynamic way users search. Users rarely view the information retrieval task as one of successively narrowing down a set of retrieved documents until a perfect match is found for some original information goal. Instead, goals change as the search proceeds: results can refine the original goal and trigger off tangential searches.

Search Visualisation using Large Screen Devices

Information visualisation is a well-established research area [7]. Much work has been put into the use of highly graphically sophisticated approaches to help the user make sense of large sets of information. Such graphical schemes have been applied to the fields of information retrieval and exploration in an attempt to overcome search problems on conventional displays. For instance, the Information Visualiser [6] allows users to manipulate an animated 3-D categorical view of search results.

These types of visualisation scheme may not be appropriate for small screen devices. Even if the display technology can deliver the high resolution required, the available screen space is not necessarily adequate for meaningful presentations and manipulation by the user. Adaptations of certain of approaches may, though, be possible [8].

Visualisation schemes that are not graphically highly intensive have also been proposed for large screen devices. An example is the Scatter/Gather approach. Similar documents are automatically clustered together and key term summaries can be displayed for each cluster. By scanning the cluster descriptions, users are able to gain an understanding of the topics available. The approach has been applied to search result output and small studies indicate it may improve users' effectiveness [15]. Schemes like the Scatter/Gather system may bring gains in the small screen context as a significant amount of information about query results can be displayed in a small space.

Search Interfaces for Small Screen Devices

We have developed a new search interface that uses the WebTwig tree outlining technique; a screenshot is shown in Figure 1. The rationale behind the scheme is that the outline view not only limits the amount of scrolling required to make sense of the search results but provides context information which should help users make decisions about which alternatives to pursue.

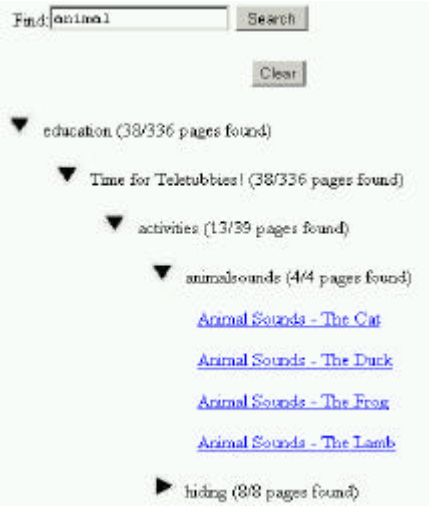

Fig. 1. WebTwig prototype search interface for handheld browser. User entered search term "animal". User was then presented with list of top-level nodes that contained hits (e.g., 'education'). By expanding nodes that contain hits, the user can progressively gain information on the context of hit. The prototype also allows the full path to a hit to be shown with one click

The PowerBrowser uses a similar approach to WebTwig for browsing, but a very different one for cross-site searching [3]. With each new search keyword, the user is

shown the number of pages in the Web site that contain the term(s). Individual page details are only shown when the user feels the number of pages in the retrieval set is small enough to deal with on the small space of the screen. The danger, of course, is that relevant and important pages may be overlooked while the user focuses on reducing the number of pages retrieved. Unfortunately, the published papers on their novel schema do not evaluate the impact of this mode of interaction, so its impact, whether beneficial or negative, is not known.

3 Experiment: Evaluation of Small Screen Searching

Given the many unanswered questions about search on the small screen given above, we decided to embark upon an evaluation that permitted some of the outstanding questions to be investigated further. As already mentioned, we wanted to compare the performance of users using a traditional web search engine on different sized screen displays.

Google is one the Web's most comprehensive search engines with over 2 billion Web pages. Although most people experience Google using a desktop computer, the company has recently started to introduce services for small screen devices.

Google on the large screen is viewed as useful and usable. Our aim here is to compare that success with user experiences when the search engine is used in both the WAP and PDA type contexts. The services are easily accessible by other workers should others wish to replicate our work.

The sorts of question we wanted to answer included: *does the small screen environment reduce users effectiveness and, if so, how?*; and, *do users alter their searching behaviour when using small screen devices?*

3.1 The Three Interfaces

WAP Interface

Figure 2 shows the interface designed by Google for mobile phones. Users enter search terms using the reduced keypad. This can sometimes be a lengthy process – most text items require multiple key pushes. In the example shown here, 56 key presses were needed to enter "mobile hci 2002" (mainly due to the large number of key pushes required for digits). Some phones employ predictive text entry schemes such as the Tegic T9™[1] system to speed up data input.

When the user presses the *search* key, the request is then handled by the Google server in the same way as for desktop queries using the entire Google index (not some subset restricted to WAP pages). The first five hits are returned to the mobile phone (see Fig. 2b) and to view further hits the user can scroll past the last result and select a *Next 5…* link.

Unlike the standard Google interface, because of the very limited screen real estate, only the title and URL is displayed for each result. WAP's horizontal scroll feature is used to display this information: for example, Fig 2b shows the first part of first

[1] http://www.t9.com/

search result's title ("MobileHCI 2002 (gio"); this is replaced with the remaining portion of the URL after a short delay (see Fig 2c).

Fig. 2. (a,b,c,d : left to right): Google on WAP type device

When a user selects a search result link, the relevant page is delivered to the device. Less than 1% of the Web's content is marked up specifically for WAP devices using WML (WAP Markup Language)[2]. Standard HTML pages cannot be displayed, so Google pre-processes each page returned. As well as basic reformatting and coding, large HTML pages are broken down into several smaller linked WML pages to fit the micro-browsers "page" (deck) size limit (around 1400 bytes). Figure 2d shows the last few lines of the first page of the "MobileHCI 2002" site; users can access further pages by selecting the *More…* link.

Previous research suggests ways of improving the WAP interface described above. First, work on news headline search result display [1] indicated that users completed search result selection tasks quicker where the full details were shown as wrapped text instead of being horizontally scrolled. Second, the basic splitting of HTML pages could be enhanced using summarisation and outlining techniques [4][5].

The PDA Type Interface

The PDA type Google interface is designed for small pocket/PDA sized computers. Figure 3 shows an example interaction. First (Fig. 3a), the user enters search terms using the device's input mechanism that might involve stylus-based handwriting or an on-screen full keyboard, for instance. Then (Fig. 3b), five search results are displayed at a time. Here, more information is provided than with the WAP scheme: the document summary and a link (*sp*) to locate similar pages. When a page is selected (Fig. 3c) the full Web page is accessible, with the user having to scroll possibly both horizontally and vertically to view information. Google does no pre-processing of the selected documents.

[2] http://www.google.com/wireless/link_wap.html

Fig. 3. (a, b,c left-to-right) PDA-size Interface for Google

Earlier work [11] showed that accessing pages designed for large screen devices on small display areas could lead to significant user problems. The search service, then, might be greatly improved if Google carried out some pre-processing on returned pages. Again, the summarisation and outlining techniques discussed in [4][5] are of interest.

Conventional Interface

The standard desktop size Google interface is well known and is widely considered usable and useful. By default double (ten) the number of results is displayed on each page than in the WAP or PDA case. Additional metadata (description and category) and facilities (e.g. to view cached copy) are also presented.

3.2 Experimental Evaluations

A controlled experiment was carried out using the three interfaces, a set of volunteer users and a range of information retrieval tasks. The aim of the experiment was to gauge the effectiveness of users using three different screen sizes (micro/WAP, small/PDA and large/conventional), and through this identify the impact on screen size on performance.

Apparatus

All 3 interfaces were presented using a conventional desktop computer and appropriate emulators/ screen sizes. All data entry was done using the standard desktop keyboard and mouse.

Use of the desktop platform allowed us to focus on the display-based interface differences. This was important for two reasons. First, it removed effects that might arise from the differing physical form factors, data entry and network conditions of the three platforms. Second, whereas the input approaches vary widely in actual devices (e.g., PDA users might use an on-screen keyboard, stylised text entry or cursive script), the display characteristics are more standardised. Search service

designers, then, will more easily gain benefits by targeting screen size issues. To reduce bias against the WAP interface, users were also not able to use the advanced search engine features available via the PDA and full screen interfaces.

For the WAP interface, the Openwave™ emulator[3] was used (see Figure 2, above). For the PDA (Figure 3) and conventional interfaces we built a browser that allowed us to fix the screen dimensions appropriately. The Microsoft IE WebBrowser2 library was used. This technology is also used in the PocketPC versions of Internet Explorer. For the PDA display version, the sizes of interactive widgets were reduced to a size consistent with those on PDA-sized devices. The usual basic range of navigational tools (forward, back, home, etc.) was provided. However, the ability of the user to use many windows, and certain other functions, were removed to avoid accidental use of multiple displays and other anomalous features in the test environment. Furthermore, the browser was extended with its own, client-side, logging system to track the progress and interaction of the user accurately.

Subjects and Tasks

We recruited 12 volunteer subjects for our experiments from within the staff and student population at a university. Before taking part, volunteers were asked to provide some information so we could assess their prior use of the Web, search engines and mobile phone and PDA technology. Most of the subjects were not computer scientists, however most saw themselves as Web 'experts'. All had used a variety of search engines (including Yahoo!, Alta Vista and Lycos) and only 3 had never used Google before. Almost all (10/12) owned a mobile phone (with the average time of ownership over 28 months), 7/11 had previously used SMS services and 5 owned a PDA.

Three realistic information retrieval scenarios were used in the experiment. Each scenario involved the user being in a city (London, San Francisco and Venice) and required them to complete 3 tourist type tasks. For example, in the London scenario the users had to find the opening hours of the National Gallery, a train schedule for London to Cambridge trains and the weather forecast. For each scenario the tasks were chosen to as similarly challenging as possible.

Experimental Method

Each user attempted to complete all 3 scenarios (that is 9 tasks). The scenarios were always presented in the order: London, San Francisco and Venice. The interface presentation order was varied for each user: 4 users were given the WAP, then the PDA and finally the conventional interface; 4 saw the conventional first, the PDA second and the WAP last; and 4 saw the PDA first, the conventional second and WAP last. Thus, each scenario was used with four users on each interface. This balanced ordering meant that all interfaces were used equally with all task scenarios and any learning or task-interface biases were reduced. In order to reduce any performance influence due to familiarity and experience, users were given a training session with each of the interface schemes immediately before attempting to carry out the tasks.

[3] Downloadable from http://www.openwave.com/products/developer_products/sdk/index.html

An observer sat next to the user during the trials and all sessions recorded on videotape. As the user carried out the tasks, they were free to verbalise their thoughts and the observer noted comments.

For each task, the time taken to complete a task and the number of tasks successfully completed was recorded. For each task, the number of search query attempts, the number of search results the user selected to look at and the number of Google search result pages examined were also recorded. After using each interface, users were asked to fill in a questionnaire to give their opinions on its usability and usefulness.

3.3 Results

The quantitative data we gathered related to users' ability to complete tasks with each interface and the interactions required for their performance. Performance was measured in time to complete the tasks and the number of tasks actually completed successfully. A task was completed either by a user giving an answer or when the user decided to stop looking for the information. A task was viewed as successfully completed if the user found information on a Web page that answered the task question. The interaction data relates to the logs of search engine user actions. Tables 1 and 2 present the data for task performance and search engine interactions for the three interfaces:

Table 1: Task performance using the different interfaces. Performance measured in time to complete (successfully or unsuccessfully) and number of tasks completed with correct answer

Interface	Time to Complete Task		Success (of 36 tasks)
	Mean (secs.)	Std. Dev.	
WAP	318	217	13 (36%)
PDA	207	157	29 (81%)
Desktop	165	160	34 (94%)

Table 2: mean search engine interactions per task. *Search attempts* is number of individual queries used in task, *search results selected* gives number of Web pages viewed as result of searches and *Google results pages viewed* indicates number of search results scanned by users (there were 5 results per page on WAP and PDA interface and 10 on conventional).

Interface	Search attempts	Search results selected	Google results pages viewed
WAP	1.8	2.1	2.9
PDA	1.6	1.5	2.0
Conventional	1.8	1.7	1.9

After using each interface, users were asked to rate search result information presented by Google in terms of the quantity and quality of information. Quantity was rated on a scale of 1 (too little) to 7 (too much); rating 4 on the questionnaire was specified as "good". Quality was rated on a scale of 1 (poor) to 7 (good). Table 3 shows the results of these ratings.

Table 3: mean rating by users of Google search result information for the three interfaces

Interface	Mean rating of information *quantity*	Mean rating of information *quality*
WAP	3.4	3.75
PDA	4.4	4.8
Conventional	4.5	4.9

Users were also asked to order four pre-specified factors in terms of how helpful they were in assisting them to decide whether to select a search result for viewing. Table 4 shows the modal ordering given by users for each of the interfaces. Finally, the subjects were asked to order six factors in terms of their negative impact on the use of their search engine. The modal ordering for each interface is shown in Table 5.

Table 4: modal ranking of factor most helpful to users when interacting with search results

WAP interface	PDA interface	Conventional interface
1. First few words of search result (the title) 2. URL of search result 3. Summary text of search result 4. Position of search result in list	1. First few words of search result (the title) 2. URL 3. Summary text of search result 4. Position of search result in list	1. Summary text of search result 2. First few words of search result (the title) 3. URL 4. Position of search result in list

Table 5: modal ranking of factors affecting users most *adversely* when interacting with search results

WAP	PDA	Conventional
1. Screen size 2. Navigation facilities 3. Search result description 4. Text/data entry facility 5. Responsiveness 6. Colours used in display	1. Screen size 2. Navigation facilities 3. =Search result description 3. =Text/data entry facility 4. Responsiveness 5. Colours used in the display	1. =Navigation facilities 1. =Search result descriptions 2. Responsiveness 3. Text/data entry facilities 4. Screen size 5. Colours used in display

3.4 Discussion

A striking result is the very poor performance of users when they used the WAP interface. Users took almost twice as long on average to succeed or give-up than when the conventional large-screen interface was used. They were also almost 60% less successful in completing tasks than the conventional case. The mean time to complete was compared with the PDA and conventional interfaces using the analysis of variance test (ANOVA) and found to be statistically significant at the 5% level ($p=0.001$). One possible explanation would be that users were obstructed by their unfamiliarity with the WAP interface. However, as stated earlier, users had been trained with each interface prior to the test, and examination of the video and transcript data gives little evidence to support such a hypothesis; there were very few instances where users could be observed or expressed having problems with the interface operation, and these were not particularly present within the WAP interface.

The PDA interface users failed to complete 14% less tasks than large screen users. This is *encouraging*: a large proportion of tasks were completed by users with the small screen interface. In our earlier reported work [11], the performance difference was much bigger with small screen users failing 50% more often. In that work, though, users mainly browsed rather than directly accessed through search. We concluded in that study that direct systematic search for small screen contexts would lead to improved usability. The results here help to validate this observation.

In all three interfaces, the numbers of search engine interface actions is small. This is consistent with other studies that show that show for instance that users usually only make one search query and rarely go beyond the second results page [10]. Interestingly, even though both the PDA and WAP interfaces display only half (five) the number of results on the first (and subsequent pages) than the large screen display, the number of results pages viewed in all cases is within the 2 to 3 range. So on average, users base their search result selection on 20 possible choices on the conventional interface and only 10 on the PDA/WAP case.

It should be emphasised that for the WAP and PDA cases, the performance we noted suggests upper limits of performance. In real-life use, WAP and PDA users will have to enter search terms using a much impoverished input device (numeric keypad and handwriting for instance) and will have to navigate with a less sophisticated tool than the conventional PC mouse.

Although there is a trend of improved performance (both time and successful completion rate) as the screen size increases (WAP-to-PDA-to-Conventional), on testing the mean time to complete tasks for PDA versus conventional screens we found no statistical significance. The reason for this lack of significance is the very high variability in completion times. We investigated these large variances further by looking at two groups of data for each interface: performance when users successfully completed a task versus that when they failed to complete. The results for each interface are shown in Table 6 (over).

For both the WAP and PDA cases, users spend over twice as long on a failed task and then give up than when they succeed. The differences in mean time of success and failure cases in WAP and PDA contexts are statistically significant. The distinction is even greater in the conventional interface case but there are only 2 failure cases and 34 successes.

With all three interfaces, then, when users succeed in completing a task, they do so quickly (within 2 to 3 minutes) and with few interactions with the search engine. When they fail, though, they fail badly.

We reviewed the logs we made during the studies and found some explanations for these two distinct patterns of use – quick successes and prolonged failures. First, we noted that problem cases were not "outlier" effects in the WAP and PDA contexts: 8 users with 9 separate tasks types accounted for the WAP failures, 5 users (and 4 different tasks) for the PDA. In the conventional interface, only 2 users (with 2 different tasks) failed.

Table 6: Task performance on WAP, PDA and conventional (Conv.) interfaces for successfully completed tasks (answer provided) and failed cases (user gave up). Levels of search engine user interactions (search attempts etc) also are shown

Interface	Outcome	Mean time to complete	Std. Dev. of completion time	Number of search attempts	Number of results selected	Number of Google pages viewed
WAP	Success	192	128	1.4	1.8	2.0
	Failure	430	222	2.2	2.4	3.7
PDA	Success	165	135	1.3	1.4	1.7
	Failure	381	117	2.6	2.3	3.1
Conv.	Success	137	113	1.7	1.5	1.8
	Failure	627	122	3.5	4.0	2.5

Exploring a search result on the WAP and PDA interfaces can involve a very high user cost in terms of time and effort. As Figures 2 and 3 illustrate, finding information within a conventional HTML page, which is being redisplayed on the smaller interface, can be a tedious, time consuming and frustrating task.

When users failed using the WAP and PDA interfaces, the main reason for failure, and the associated large task timings, was the great difficulties they had in navigating the site selected from the search result. Most of their wasted time and effort was spent in becoming increasingly lost within the small window. As Table 6 indicates, failing users also carried out more search engine interactions: they carried out a greater number of search attempts, browsed more of the search result pages and selected more of the search results.

The impression when observing these cases was of users 'thrashing' to try and solve the problem. They would carry out an initial search attempt, spend more time scanning the search result outputs, explore a search result and become lost and frustrated, then return to the search engine for another fruitless attempt.

In the unsuccessful cases, often it seems that users were very uncertain about whether a search result they were about to explore was going to be of any use. They then made blind leaps of faith into a usually disappointing unknown.

The successful cases for the WAP and PDA contexts were where the search engine results contained "obviously" good candidates. These results were the ones where even the limited information about the page (title and URL for WAP and title, URL and limited summary for PDA) was enough to suggest the page was worth exploring.

In real-world use (using the physical devices rather than emulators) we might expect to see more of the unsuccessful cases – as search term entry is expensive on the impoverished interface, less expressive queries might be entered, leading to poorer search results. Furthermore, users might be less inclined to review search lists due to the navigation costs.

User Subjective Ratings

As well as differences in the measurable performance, as the screen size increases, so do users' satisfaction with the quantity and quality of information provided by Google (see Table 3). However, the range in ratings, around one point from lowest to highest, is not as wide as might be expected. A possible explanation for this is that Google have produced a very simple, uncluttered interface for all three devices.

Table 4 indicates users see the more descriptive elements of search results as most important. Titles are favoured over URLs and where detailed summary text is available (the conventional interface case) this is rated the most important.

The differences in users' views about factors that adversely affected their behaviour (see Table 5), show that screen size was seen as biggest limitation for WAP and PDA (it is rated fourth out of fifth in the conventional case). Leaving aside screen space, for all three interfaces, limitations of the search engines navigation facilities (manipulating the result sets) and search result information were rated as negatively affecting user behaviour.

4 Design Guidelines for Small Screen Search Engine Interfaces

Clearly screen size has a major impact on user performance. Success rates drop and even the time to complete successful searches increases. From our evaluations and observations, we propose several ways that the Google WAP and PDA interfaces might be improved. These guidelines will also be of interest to others developing search interfaces for small screen contexts.

1. Reduce the amount of page-to-page navigation needed to view search results. Users do not look at many search result pages and also prefer not to shuffle with groups of pages to view information. As we observed in [11], page-to-page navigation is very costly when browsing in general, and in our current observations we have seen similar behaviour when users are browsing search results. Although increasing the number of results on a WAP card or PDA screen will lead to increased vertical scrolling, this additional user effort affects performance to a lesser extent than the page-to-page navigation.

2. Provide more rather than less information for each search result. Users value good quality information about search results. As we have seen, selecting a search result, particularly for WAP, is a very "risky" action. Users were clearly observed seeking information to guide their next step as they browsed the search result list, and expressed uncertainty when given what they felt was inadequate information. Better quality information should support user confidence, and if appropriate should also enhance performance. For the WAP interface especially, more

information should be provided and should be presented using the wrapped round text rather than the automatic horizontal scroll method [1]. Clearly, for WAP given the limited deck size, there needs to be a technical trade-off between this guideline and the first.

3. Provide a quick way for users to know whether a search result points to a conventional HTML page or a small screen optimised page. If search results are not optimised for WAP pages, there is very little point in WAP users selecting them: users will simply become lost as they struggle through the many WAP cards needed to represent the HTML page. We observed that frames-based sites can be particularly damaging, even with sophisticated conversion. Although the larger display area on PDA type computers reduces the problem, pages adapted for these devices will be easier to use. The search result list could use a small icon or text device to let users scan and find small screen suitable information. It may not be possible, for any one of a host of reasons, to provide a small-screen optimised version of a site. Where an optimised form is available, users will generally perform better, through reduced scrolling, so assist them in making an informed choice before committing (and then often failing on poorly converted pages).

4. Pre-process conventional pages for better usability in small screen contexts. Google already pre-processes non-WAP pages so they can be displayed on WAP devices. More sophisticated adaptations for both WAP and PDA sized screens are possible (see Section 3.1). This could lead to much increased user effectiveness.

5. Adapt for vertical scrolling – in our first evaluation [11] and our observations in this evaluation, users tend to scroll *vertically* rather than *horizontally* – design with this bias in mind; information which requires significant sideways scrolling will often never be seen.

5 Conclusions and Future Work

More and more people will soon be using small, handheld devices to search the Web. Anticipating this, Google (along with other search providers) have begun to introduce services for these new platforms. Although the technology has been 'optimised' for the *device* capabilities (bandwidth, memory sizes etc), our work suggests that further optimising for *user* capabilities would improve such services greatly. User-based experiments are important, particularly as mobile Internet devices are developing quickly without a transparent process for careful user-centred design.

As screen size is reduced, from full screen to PDA-sized and yet further to mobile phone dimensions, user performance drops. The main reason for this is that on a smaller screen it becomes increasingly more difficult for a user to make good quality judgements about the usefulness of any particular search result. Poor search result choices can be disastrous in human-computer interaction terms: some of our users became completely lost, spending 10 minutes trying to find information on a WAP screen that took 10 seconds to locate on a conventional desktop computer.

Using the search engine via a WAP phone was very ineffective. However, the performance on the PDA-sized screen was encouraging. Our previous studies suggested that for this sort of screen area, user performance would be good if direct,

search-based access was provided to Web resources; the results here add weight to this claim.

This study focussed on the screen-size issues. Now that we have "upper-limit" indications on performance, we are extending our evaluations to look in more detail at interaction problems when the search services are used on the physical devices. We expect the patterns seen in this work to be repeated but are interested in measuring, for example, search query length used given different text entry mechanisms.

For improvements in both WAP and PDA-sized devices, search engine designers need to develop interaction schemes that allow users to better assess search results. Users should be able to make good choices quickly. Further, when a conventional Web page is re-displayed on the smaller devices pre-processing is needed to help users navigate within the information. We are using the results of this study to further develop and evaluate the WebTwig [12] and alternative approaches for small screen search.

6 Acknowledgements

Thanks to Craig Nevill-Manning of the Research Group at Google Inc. who provided technical information and advice about the WAP and PDA search index and pre-processing.

References

1. Buchanan, G., Farrant S, Jones, M., Thimbleby, H., Marsden, G. & Pazzani M. (2001). Improving mobile Internet usability. Proceedings of the Web 10 conference on World Wide Web, 673 – 680. Hong Kong, April 2001
2. Buyukkokten, O., Garcia-Molina, H., Paepcke, A. & Winograd, T. (2000). Power browser: efficient Web browsing for PDAs. Proceedings of the CHI 2000 conference on Human factors in computing systems, pp 430 – 437. Amsterdam.
3. Buyukkokten, O., Garcia-Molina, H. & Paepcke, A. (2000). Focused web searching with PDAs. Proceedings of Web 9 Conference, Amsterdam.
4. Buyukkokten, O., Garcia-Molina, H. & Paepcke, A. (2001) Accordion Summarization for End-Game Browsing on PDAs and Cellular Phones. Human-Computer Interaction Conference 2001 (CHI 2001). Seattle, Washington - 31 March-5 April, 2001
5. Buyukkokten, O., Garcia-Molina, H. & Paepcke, A.. (2001) Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices. The 10th International WWW Conference (WWW10). Hong Kong, China - May 1-5, 2001.
6. Card S.K, Robertson, G.G. & Mackinlay, J. D. (1991). The information visualizer, an information workspace. Proceedings of Human factors in computing systems (CHI 91), pp181–186
7. Card, S. K., Mackinlay, J. D. & Schneiderman, B. (1999). Readings in Information visualisation – using vision to think. Morgan Kaufmann Publishers, Inc.

8. Dunlop, M. D. & Davidson, N. (2000). Visual information seeking on PDAtop devices. Volume II Proceedings of BCS HCI 2000, pp19-20, Sunderland, September 2000.
9. Hearst, M. (1999). User interfaces and visualisation. In Modern Information Retrieval, R. Baeza-Yates & B. Ribeiro-Neto (Eds). ACM Press/Addison-Wesley.
10. Jansen, B J, Spink, A, Saracevic, T. (2000). Real life, Real Users and Real Needs: A Study and Analysis of User Queries on the Web, Information Processing and Management, 36(2), 207-227
11. Jones, M, Marsden, G., Mohd-Nasir, N, Boone, K, & Buchanan, G. (1999) Improving web interaction in small screen displays. Proceedings of Web 8 conference, pp51–59, Toronto 1999
12. Jones M, Buchanan, G & Mohd-Nasir, N (1999). Evaluation of WebTwig — a site outliner for handheld Web access. International Symposium on Handheld and Ubiquitous Computing, Karlsrhue, Germany. Gellerson, H-W (Ed.), Lecture Notes in Computer Science 1707:343–345, Springer-Verlag.
13. Ling R., editor, *Personal and Ubiquitous Computing*, Special Issue on Mobile Communication, **5**(2), 2001.
14. Marsden, G., Gillary, P., Thimbleby, H. & Jones. M. The Use of Algorithms in Interface Design. International Journal of Personal and Ubiquitous Technologies. Springer-Verlag Volume 6 issue 1, January (2002)
15. Pirolli, P., Schank, P., Hearst, M. & Diehl, C. (1996). Scatter/ Gather browsing communicates the topic structure of a very large text collection. Proceedings of Human factors in computing systems conference (CHI'96):213–220.
16. Shneiderman, B., Byrd, D. & Croft, B (1998). Sorting out searching. Communications of the ACM, 41(4):95–98.