



Available at  
[www.ComputerScienceWeb.com](http://www.ComputerScienceWeb.com)  
POWERED BY SCIENCE @ DIRECT®

**Interacting  
with  
Computers**

Interacting with Computers 15 (2003) 479–495

[www.elsevier.com/locate/intcom](http://www.elsevier.com/locate/intcom)

## Improving web search on small screen devices

Matt Jones<sup>a</sup>, George Buchanan<sup>b</sup>, Harold Thimbleby<sup>c,\*</sup>

<sup>a</sup>*Department of Computer Science, University of Waikato, Hamilton, Waikato, New Zealand*

<sup>b</sup>*Interaction Design Centre Middlesex University, London, UK*

<sup>c</sup>*University College London Interaction Centre (UCLIC), 26 Bedford Way, London, UK*

---

### Abstract

Small handheld devices—mobile phones, Pocket PCs etc.—are increasingly being used to access the web. Search engines are the most used web services and are an important factor of user support. Search engine providers have begun 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 search tasks using a mobile phone-sized, handheld computer-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 2–3 min) 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. In addition, we present and discuss novel interaction schemes that put these guidelines into practice.

© 2003 Elsevier B.V. All rights reserved.

*Keywords:* Search; Interfaces; Evaluation; Design; Mobile computing

---

### 1. Introduction

Research into usable, useful and effective interactions for web search is vital. Unless effective user-centered approaches are developed and applied, the promise of effective access to information resources will be lost, with users left frustrated and overwhelmed (Shneiderman et al., 1998). Small screen access presents special challenges to interaction designers—unless interfaces take special account of the reduced screen space, users' performance and satisfaction are likely to be much reduced.

Recently, Google™, one of the most comprehensive web search engines, began to offer its services to wireless access protocol (WAP) phone users and via personal digital

---

\* Corresponding author.

*E-mail addresses:* [h.thimbleby@cs.ucl.ac.uk](mailto:h.thimbleby@cs.ucl.ac.uk) (H. Thimbleby), [always@acm.org](mailto:always@acm.org) (M. Jones), [george10@mdx.ac.uk](mailto:george10@mdx.ac.uk) (G. Buchanan).

assistant (PDA) type handheld computers. We evaluated the usability of these services to explore the impact of screen size on user performance.

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, 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 small screen contexts. We have also developed some experimental search interfaces for small screen devices that put these guidelines into practice.

The paper is organized as follows. In Section 2, we review existing work on search interfaces and mobile access. The experimental evaluation is presented in Section 3 and details of how we have used the findings of the experiment are given in Section 4. Section 5 concludes the work.

## **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 small display areas. The results of one study (Jones et al., 1999a) suggested that users did not want to use conventional page-to-page navigation as it was interactively very costly. Rather, a much more direct, systematic approach requiring less scrolling was seen as appropriate.

WebTwig (Jones et al., 1999b) was developed to demonstrate a directed approach to handheld browsing that takes account of the limited display. The tool uses an outline-style interaction technique. A hierarchical view of the web site is presented that can be interactively expanded or contracted. The aim is to enable users to identify useful areas of the site before a final document selection is made. In a user evaluation that compared WebTwig with the standard, non-adapted view, the mean task completion time fell by 35% and success rates rose. A similar scheme for the PalmPilot has also demonstrated the efficacy of outlining presentation and interaction in handheld browsing (Buyukkokten et al., 2000a).

### *2.2. Interfaces for searching*

Most search systems simply present the results of a user query as a (long) ranked list of matching documents, often broken into pages. With such an interface, users have to scroll and page through the often-overwhelming list, examining descriptions and perhaps document content in detail as they proceed to make relevance judgements. Such approaches mean that even on conventional large displays, search interfaces have usability problems.

Hearst (1999) 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 user's search proceeds: results can refine their original goal and trigger off tangential searches, and as the cognitive load on the user increases, they are more likely to make errors and never return to their original goal (which may not be a problem).

### *2.3. Search visualisation using large screen devices*

Information visualisation is a well-established research area (Card et al., 1999). Much work has been put into the use of highly graphical 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 (Card et al., 1991) allows users to manipulate an animated 3D 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 approaches may, though, be possible—for example Dunlop and Davidson (2000) have experimented with Starfield querying (Alhberg and Schneiderman, 1994) on a PalmPilot.

Visualisation schemes that are not graphically highly intensive have been proposed for large screen devices. One 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 (Pirolli et al., 1996). 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.

Despite these sorts of research innovation, list or list-like presentations still predominate on the large screen. There seems to be a 'paper-isation' of search interfaces: preserving a particular approach long after the technological necessities that generated it have been removed.

### *2.4. Search interfaces for small screen devices*

Small screen search interfaces made available by the major commercial search engines all use the conventional list presentation. Some account is made of the limited screen space by, for example, reducing the number of results presented in each result page and limiting the amount of information displayed for each result.

On the other hand, the PowerBrowser (Buyukkokten et al., 2000b) includes a search interaction scheme that is more clearly adapted for the small screen space of a PDA. 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 scheme do not evaluate the impact of this mode of interaction, so its influence, whether beneficial or negative, is not known. A related, evaluated approach, applied to mobile phone function access, is reported in Marsden et al. (2002). In their scheme, as users ‘spell out’ function names letter-by-letter, potential matches are displayed; this list is dynamically updated as more of the name is keyed in. This scheme was compared with the conventional menu based approach and found to be more effective.

Both PowerBrowser and our own WebTwig system bear similarities to the WebTOC system of Nation et al. (1997). WebTOC provides an outline-style topic-centred access to large hypertext (i.e. web) structures, which demonstrated improved performance on the desktop, a result echoed in our own findings with WebTwig on smaller displays. However, WebTOC used a human-generated classification scheme, as opposed to the automatic systems of WebTwig and PowerBrowser

As well as adapting the information structure and presentation for small screen viewing, researchers have examined other ways to improve mobile search interaction. For example, there are adaptive search engines that personalise results (Billsus et al., 2002). Use of context information about the location of a mobile device is also seen as important in providing richly appropriate assistance to users (e.g. Kießling and Balke, 2002).

Aridor et al. (2002) tackle the overload problem with a ‘focussed search’ approach that combines both online and offline access. Users define a number of ‘knowledge-agent bases’ for topics of interest. A conventional PC system then attempts, over time, to learn more about that topic and then extracts related web pages that are transferred to the handheld computer. These sets of information can be searched offline and the key pages can be viewed in full whilst less relevant pages can be accessed if the user has a wireless connection to the web.

So, in summary, then, focussed approaches to browsing and searching provide a consistent means of improving information seeking and interaction on small-screen devices. Simple, outline-style methods of presentation can be complemented with aspects of personalisation and other context- or use-sensitive techniques to direct user attention onto salient elements.

### 3. Experiment: evaluation of small screen searching

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 relevant 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?*

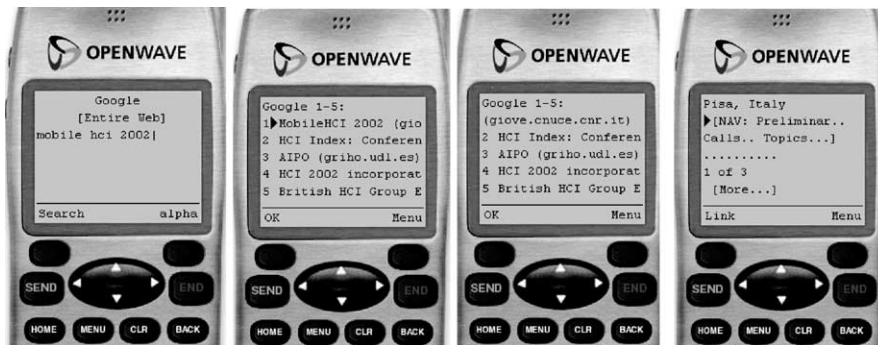


Fig. 1. (a–d: left to right): Google on a WAP type device.

### 3.1. The three interfaces

#### 3.1.1. WAP Interface

Fig. 1 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™ system<sup>1</sup> to speed up data input.

When the user presses the *search* key, the first five hits are returned to the mobile phone (see Fig. 1(b)) 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. 1(b) shows the first part of first search result’s title (‘MobileHCI 2002 (gio)’); this is replaced with the remaining portion of the URL after a short delay (see Fig. 1(c)).

When a user selects a search result link, the relevant page is delivered to the device. Only around 0.2% of the web’s content is marked up specifically for WAP devices using WML (WAP markup language)<sup>2</sup>. Standard, non-WAP, HTML pages are therefore reformatted and re-coded. Large HTML pages are broken down into several smaller, linked WML pages to fit the micro-browsers ‘page’ (deck) size limit (around 1400 bytes). Fig. 1(d) shows the last few lines of the first page of the ‘MobileHCI 2002’ site; users can access further pages by selecting the *More...* link.

#### 3.1.2. The PDA type interface

The PDA type Google interface is designed for small pocket/PDA sized computers. Fig. 2 shows an example interaction. First (Fig. 2(a)), the user enters search terms using the device’s input mechanism that might involve stylus-based handwriting or an on screen full

<sup>1</sup> <http://www.t9.com/>

<sup>2</sup> [http://www.google.com/wireless/link\\_wap.html](http://www.google.com/wireless/link_wap.html)



Fig. 2. (a–c left-to-right) PDA-size user interface for Google.

keyboard, for instance. Then (Fig. 2(b)), five search results are displayed at a time. When a page is selected (Fig. 2(c)) 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.

### 3.1.3. Conventional interface

The standard desktop size Google interface is well known and is generally considered usable and useful (or at least its usability is not an issue in the context of the present paper). By default double the number of results (i.e. ten) is displayed on each page compared to 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.

### 3.2.1. Apparatus

All three interfaces were presented using a conventional desktop computer and appropriate emulators and 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.

For the WAP interface, the Openwave™ emulator<sup>3</sup> was used (see Fig. 1). For the PDA (Fig. 2) and conventional interfaces we built a browser that allowed us to fix the screen dimensions appropriately. 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.

### 3.2.2. *Subjects and tasks*

We recruited twelve volunteer subjects for our experiments from within the staff and student population at a university. Before taking part, volunteers were asked to provide 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’.

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 three tourist type tasks—one concerned with a tourist attraction, another relating to transport and the third to either a hotel listing or weather report. 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 be as similarly challenging as possible: an expert user was asked to perform the tasks on a conventional desktop machine and the number of interactions and time taken to complete the tasks was measured to assess the relative difficulty of the three scenarios.

### 3.2.3. *Experimental method*

Each user attempted to complete all three scenarios (that is nine tasks). The scenarios were always presented in the order: London, San Francisco and Venice. The interface presentation order was varied for each user: four users were given the WAP, then the PDA and finally the conventional interface; four saw the conventional first, the WAP second and the PDA last; and four 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 that 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.

An observer sat next to the user during the trials and all sessions were recorded on videotape. As the user carried out the tasks, they were free to verbalise their thoughts and the observer noted their comments. For each task, the time taken to complete a task and the number of tasks successfully completed was recorded, though it should be noted that think aloud protocols make users take longer, and therefore reduce the reliability, in particular,

<sup>3</sup> Downloadable from [http://www.openwave.com/products/developer\\_products/sdk/index.html](http://www.openwave.com/products/developer_products/sdk/index.html)

of absolute timings, and to a lesser extent time comparisons. 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.

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.

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](#).

### 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. Users were also almost 60% less successful in completing tasks than in the conventional case. The WAP mean time to complete was compared with the PDA and conventional interfaces using the analysis of variance test (ANOVA) and is statistically significant at the 5% level ( $p = 0.001$ ).

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 (s)	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* are 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 five results per page on WAP and PDA interface and ten 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
Desktop	1.8	1.7	1.9

The PDA interface task completion performance was 85% of that seen in the large screen case. This is encouraging to us: a large proportion of tasks were completed by users with the small screen interface. In our earlier reported work (Jones et al., 1999a), the performance difference was much bigger with small screen users failing 50% more often than large screen ones. In that work, though, users completed tasks by mainly browsing rather than using search. We concluded in that study that direct systematic search for small screen contexts would lead to improved usability. The results here help to support this observation.

In all three interfaces, the numbers of search engine interface actions is small. This is consistent with other studies that show, for instance, that users usually make only one search query and rarely go beyond the second results page (Jansen et al., 2000). 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 2–3 range. So on average, users base their search result selection on 20 possible choices on the conventional interface and only 10–15 on the PDA/WAP case. Such transfer effects from the large screen environment present interesting challenges to small screen interaction designers.

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

Table 3

Mean rating by users of Google search result information for the three interfaces

Interface	Mean rating of information <i>quantity</i>	Mean rating of information <i>quality</i>
WAP	3.4	3.8
PDA	4.4	4.8
Desktop	4.5	4.9

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

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

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.

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 two failure cases and thirty four successes.

With all three interfaces, then, when users succeed in completing a task, they do so quickly (within 2–3 min) 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.

Exploring a search result on the WAP and PDA interfaces can involve a very high user cost in terms of time and effort. As Figs. 1 and 2 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

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

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

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

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.

#### 3.4.1. User subjective ratings

As the screen size increases, as well as differences in the measurable performance, there are changes in 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 that users perceive the more descriptive elements of search results as most important in helping them discriminate between search results. 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. In contrast, it is rated fourth out of five in the conventional case. Leaving aside screen space,

for all three interfaces, limitations of the search engines navigation facilities (manipulating the result sets) and the limited search result information were rated as negatively affecting user behaviour.

#### 4. Acting on experimental findings

An important role of experimentation in HCI should be to spur improvements in, and innovations of, existing schemes and systems. We have built on our experimental findings in two ways: we have developed a set of design guidelines for small screen search interfaces; and, used these to help develop two novel small screen search systems.

##### 4.1. Design guidelines

Clearly screen size has an 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 small screen search interfaces might be improved.

*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 (Jones et al., 1999a) 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.

*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 (Buchanan et al., 2001). For WAP, given the limited deck size, there needs to be a technical trade-off between this guideline and the first.

*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).

*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 (e.g. Buyukkokten et al., 2001a,b). This could lead to much increased user effectiveness.

*Adapt for vertical scrolling*—in our first evaluation (Jones et al., 1999a) 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.

Although we have arrived at these guidelines through studying an online, general mobile search engine, they can be applied to more specialised mobile search schemes such as those that employ contextual filters to the search results (see Section 2.2).

#### 4.2. Novel small screen search interfaces

We have developed two small screen search interfaces, LibTwig, an outline-view search tool, and a key phrase based system.

##### 4.2.1. Outline-view search

LibTwig extends the WebTwig browsing tool to allow searching. Two search result presentations are possible (see Fig. 3). The first is the conventional relevance ranked list. The second is an outline view that presents search result documents in relation to where they occur in the information hierarchy. For example, in Fig. 3 (centre), on querying with ‘snail,’ the user is shown there are several top-level categories (including ‘Agriculture and Food Processing’ and ‘Animal Husbandry and Animal Product Processing’) that contain ‘snail’ documents. On expanding the first category, the user sees that two such documents are within the ‘Better Farming series of FAO and INADES’ sub-category. Finally, on



Fig. 3. LibTwig in use. Left—conventional ranked list; Centre—outline search with two top level categories visible and first expanded to show sub-category; Right—outline search with sub-category expanded and a document link visible.

further expanding this sub-category, they are shown the document links themselves (see rightmost screenshot).

As with browsing, the outline view not only limits the amount of scrolling required to make sense of the search results but provides context information which might help users make decisions about which alternatives to pursue.

The Cluster Hypothesis (van Rijsbergen, 1979) also motivates the scheme. It predicts that most ‘real’ matches for a search will be in a common classifier, so a corollary is that matches not in that category are probably less relevant. Therefore, the thematic division of results by subject category might also improve the effective selectivity of the document review task.

The results of a first evaluation of LibTwig indicated that user performance with the outline presentation compares well with the ranked list scheme (Buchanan et al., 2002). Encouraged by these results, we are making LibTwig more sophisticated by, for example, exploiting rank information within the hierarchical presentation.

#### 4.2.2. Key phrase-based system

Small screen search interfaces should provide users with as rich information as possible to help them discriminate between and select useful search results. The WAP Google interface only presents the title and URL of each result; this is too little, particularly as many web pages are likely to have titles that poorly describe the content. The PDA Google

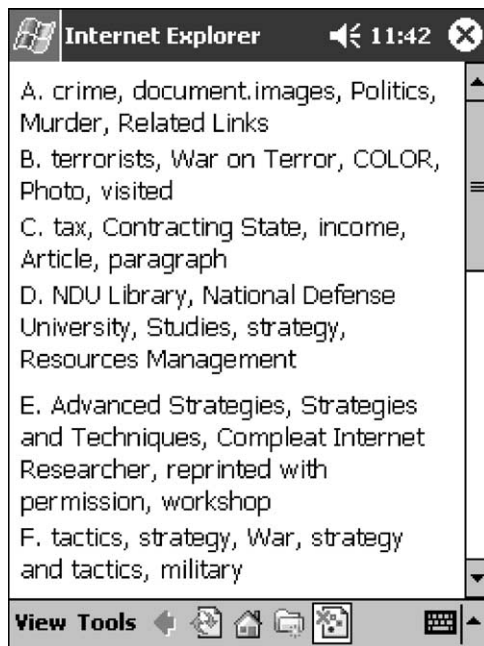


Fig. 4. Example search result list in the key phrase system as viewed on a Pocket PC. Each document (A–F) is described solely by a set of key phrases automatically extracted from those documents.

system does give small excerpts of the text showing keyword match contexts for each result.

We are examining the use of automatically extracted key phrases as useful document surrogates for small screen search result discrimination. Key phrases could present the ‘essence’ of a document while at the same time using only a small amount of screen space. An example result list (with only key words and no titles) is shown in Fig. 4. Early evaluation results indicate that key phrases alone are as useful as good titles in helping users to make decisions about the document contents (Deo, 2002). Another consideration in favour of document summarisations such as this is that given the prevalence of poor document titles, particularly in the case of web pages, such approaches can yield a more meaningful alternative descriptor for documents.

## 5. Conclusions and future work

More and more people are using small handheld devices to search the web. Following this trend, Google, along with other providers, have begun to introduce special services for these new platforms. Although the technology may be targeted for *device* capabilities (bandwidth, memory sizes etc.), our work suggests that 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 smaller screens make it increasingly more difficult for a user to make good quality judgements about the usefulness of any particular search result or to gain a general overview effectively. Poor search result choices can be disastrous in human-computer interaction terms: some of our users became completely lost, spending 10 min trying to find information on a WAP screen that took 10 s to locate on a conventional desktop computer. The knock-on consequences when the user is trying a real-life task would be worse than the performance measures of our experiments directly indicate.

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 size, user performance would be good if direct, search-based access was provided to web resources; our empirical results here add weight to this claim.

This study focussed on the screen-size issues in lab-based conditions. We are extending our evaluations to look in more detail at interaction problems when the search services are used on the physical devices in mobile contexts. We expect the broad patterns seen in this work to be repeated, but we are interested in measuring, for example, search query length used given different text entry mechanisms. As others have noted, we see the combination of laboratory-based investigations and contextual mobile studies as being important (Waterson et al., 2002). Although contextual studies are important, it is worth noting that many ergonomic, perceptual and biomechanical issues can be studied very effectively in the laboratory.

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 outline and alternative approaches for small screen search.

### 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. The work on key phrases is being carried out with Steve Jones and Shaleen Deo of the University of Waikato. Harold Thimbleby is a Royal Society-Wolfson Research Merit Award holder, and acknowledges their support.

### References

- Alhberg, C., Schneiderman, B., 1994. Visual information seeking: tight coupling of dynamic query filters with starfield displays. *Proceedings CHI94*, Boston, USA, 311–317.
- Aridor, Y., Carmel, D., Maarek, Y.S., Soffer, A., Lempel, R., 2002. Knowledge encapsulation for focused search from pervasive devices. *ACM Transactions on Information Systems* 20 (1), 25–46.
- Billsus, D., Brunk, C.A., Evans, C., Gladish, B., Pazzani, M., 2002. Adaptive interfaces for ubiquitous web access. *Communications of the ACM* 45 (5), 34–38.
- Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., Pazzani, M., 2001. Improving mobile Internet usability. *Proceedings Web 10th conference on World Wide Web*, Hong Kong, pp. 673–680.
- Buchanan, G., Jones, M., Marsden, G., 2002. Exploring small screen digital library access with the greenstone digital library. In: Agosti, M., Thanos, C. (Eds.), *Proceedings of the sixth European Conference on Research and Advanced Technology for Digital Libraries*, Rome. *Lecture Notes in Computer Science*, vol. 2458. Springer-Verlag, Berlin, pp. 583–596.
- Buyukkokten, O., Garcia-Molina, H., Paepcke, A., Winograd, T., 2000a. Power browser: efficient web browsing for PDAs. *Proceedings CHI2000*, Amsterdam, 430–437.
- Buyukkokten, O., Garcia-Molina, H., Paepcke, A., 2000b. Focused web searching with PDAs. *Proceedings Web*, Amsterdam 9, 213–230.
- Buyukkokten, O., Garcia-Molina, H., Paepcke, A., 2001a. Accordion summarization for end-game browsing on PDAs and cellular phones. *Proceedings CHI 2001*, Seattle, Washington, 213–220.
- Buyukkokten, O., Garcia-Molina, H., Paepcke, A., 2001b. Seeing the whole in parts: text summarization for web browsing on handheld devices. *Proceedings of 10th International WWW Conference (WWW10)*, Hong Kong, China.
- Card, S.K., Robertson, G.G., Mackinlay, J.D., 1991. The information visualizer, an information workspace. *Proceedings CHI 91*, 181–186.
- Card, S.K., Mackinlay, J.D., Schneiderman, B., 1999. *Readings in Information Visualisation—Using Vision to Think*, Morgan Kaufmann Publishers, Los Altos, CA.
- Deo, S., 2002. Search document access on small screen devices. MSc Thesis, University of Waikato, New Zealand.
- Dunlop, M.D., Davidson, N., 2000. Visual information seeking on PDA top devices. *Proceedings BCS HCI 2000, Sunderland II*, 19–20.
- Hearst, M., 1999. User interfaces and visualisation. In: Baeza-Yates, R., Ribeiro-Neto, B. (Eds.), *Modern Information Retrieval*, ACM Press/Addison-Wesley, Reading, MA, USA.



- 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.
- Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G., 1999a. Improving web interaction in small screen displays. *Proceedings Web 8*, 51–59.
- Jones, M., Buchanan, G., Mohd-Nasir, N., 1999b. In: Gellerson, H.-W., (Ed.), *Evaluation of WebTwig—a site outliner for handheld Web access*, *Proceedings International Symposium on Handheld and Ubiquitous Computing, Karlsruhe*. *Lecture Notes in Computer Science*, vol. 1707. Springer, Berlin, pp. 343–345.
- Kießling, W., Balke, W.-T., 2002. Mobile search in a preference world. In: Maarek, Y., Soffer, A. (Eds.), *Working Notes of the WWW2002 Workshop on Mobile Search, Honolulu, Hawaii*, pp. 32–38, <http://www.haifa.il.ibm.com/Workshops/www2002-MobileSearch/>.
- Marsden, G., Gillary, P., Thimbleby, H., Jones, M., 2002. The use of algorithms in interface design. *International Journal of Personal and Ubiquitous Technologies* 6 (2), 132–140.
- Nation, D.A., Plaisant, C., Marchionini, G., Komlodi, A., 1997. Visualizing websites using a hierarchical table of contents browser: WebTOC, *Proceedings Third Conference on Human Factors and the Web, Denver Co.*.
- Pirulli, P., Schank, P., Hearst, M., Diehl, C., 1996. Scatter/gather browsing communicates the topic structure of a very large text collection. *Proceedings CHI 1996*, 213–220.
- Shneiderman, B., Byrd, D., Croft, B., 1998. Sorting out searching. *Communications of the ACM* 41 (4), 95–98.
- van Rijsbergen, C.J., 1979. *Information Retrieval*, second ed, Butterworths, London.
- Waterson, S., Landay, J.A., Matthews, T., 2002. In the lab and out in the wild: remote web usability testing for mobile devices. *Extended Abstracts, CHI 2002*, 796–797.