Changing the Pace of Search: Supporting "Background" Information Seeking

Matt Jones[1], George Buchanan, Tzu-Chiang, Preeti Jain
Department of Computer Science,
University of Waikato, New Zealand
always@acm.org
Fax: +64 7 858 5095

ABSTRACT

Almost all Web searches are carried out while the user is sitting at a conventional desktop computer, connected to the Internet. Although online, handheld, mobile search offers new possibilities, the fast-paced, focused style of interaction may not be appropriate for all user search needs. In this paper, we explore an alterative, relaxed style for web searching that asynchronously combines an offline handheld computer and an online desktop Personal Computer. We discuss the role and utility of such an approach, present a tool to meet these user needs and discuss its relation to other systems.

Keywords

Search scenarios, interaction style, handheld device, slow technology.

[1] Author for correspondence (always@acm.org)

INTRODUCTION

While most Web searching is done sitting at a conventional, desktop PC, it is likely that search needs take shape away from the desktop, stimulated by activities a user is involved in, such as attending a meeting or listening to a presentation or lecture.

Mobile, handheld devices are beginning to be used to provide *online* search access. Clearly, these services will be useful, especially to meet specific, focused and urgent information needs. However, these approaches ask much of the user who has to engage in a cognitively demanding "foreground" information seeking process (Jones et al, 2003a). For some use-contexts and information needs, this burden is unhelpful and inappropriate. In this paper, we consider the need to support users in "background" information seeking.

BACKGROUND INFORMATION SEEKING

In an interesting diary-style study (Brown et al, 2000), the types of information people record, and how they use this information subsequently, were analysed. Participants often made notes for later use in some other task such as document preparation or as a reminder to pursue something of interest.

From this study we can envisage that people routinely make notes on memo pads, scraps of papers, diaries and the like with the intention of later use with a search engine. Unlike urgent, focused information seeking, when the information need arises (say in a seminar presentation), it does not have to be satisfied immediately; and, the process of satisfying the need is not the user's main focus (while listening to the presenter is, for instance).

In applying their findings, Brown *et al*, propose that capture tools (be they computerized or physical) should allow users to make sense of and organize their notes, and to be reminded in some way to act on them. The current approaches people use to capture their potential queries seem to fall far short of these

ideals. It is likely that many search inspirations are not acted upon: firstly, perhaps, the bits of paper are lost before the user gets back to their desktop PC; alternatively, when they return to their PC they might not be able to read the scrawling handwriting on their own notes; and, even if the notes are legible, they may not remember the significance of the proposed search.

Putting aside the problems of search query capture and later use, earlier work indicates that many potential information needs, while the user is away from their desktop, will not even be formulated. Taylor (Taylor, 1962), presented a four-stage model of the search process that describes how a user's search need moves from being "visceral" (a vague, unexpressed incompleteness in their knowledge), to "conscious" and "formalized" (an "ideal" query statement) and finally "compromised" (the actual query presented via the facilities provided by the information retrieval system). In the sorts of situation we have outlined (like attending a seminar) users may have "visceral" needs but may not move to express these needs, at the time, more formally, without some systematic support.

To explore this issue, we asked six graduate students at a University to keep a week long log, noting whenever they were aware of any "conscious" search needs (the second step of Taylor's model). Only 44 potential search terms were recorded over the 7-day period for all users. In contrast, when we asked them to estimate the number of times they had used an online search engine in the same period, the average number of sessions was thirty. The main reason our participants gave to explain the difference between the low number of expressions of information need in their diaries and the large number of online search sessions was that they viewed search as a *computer based activity*; the presence of an online search tool, it seems, motivated their search actions.

While it is important to support dynamic, fast-paced, iterative interaction with search results so that users can refine and shape their queries, we also know that people make use of the results of information seeking processes in a range of slower, ambient ways. For instance, observations of information seeking

in physical environments have highlighted the ways knowledge workers place implicit cues in their workplace for later action (Kidd, 1994).

THE LAID-BACK SEARCH TOOL

We have developed a prototype tool that seeks to provide a search style that parallels and complements the effective yet asynchronous, opportunistic and casual forms found beyond the electronic environment. The system is designed to allow users to capture search needs in the background using a handheld computer, providing a tool that has the qualities noted in (Brown et al, 2000) as well as acting as a 'prop' to assist users in turning their "visceral" needs into "conscious" ones (Taylor, 1962). Further it helps users to make use of the results of search in slower, more reflective ways than conventionally possible with online search.

Capturing search terms

The user enters search terms on an offline handheld computer. They can record simple keyword searches and use the "advanced search" interface to restrict the scope of the search to a particular web site *etc.* (Figure 1). The handheld application performs a check for duplicate entries and allows the complete set of queries to be viewed and edited.

The handheld, then, gives the user greater support for query noting than when they simply use paper-and-pen. However, as with paper-and-pen, the user's main focus – such as listening to a talk – is not greatly disrupted while their search need is captured.

An initial evaluation of the handheld system (Jones et al, 2003) showed a need for the tool that to integrate better with other handheld applications. To address this, we extended the system so it can also scan all of the user's handheld notes, calendar and task entries and extract words and phrases which have been marked by the user for possible querying. The user can select from the extracted terms and add these to any explicitly entered queries (see Figure 2).

Processing the captured queries

When the handheld is reconnected to the PC, all of the queries captured offline are sent to a search engine (Google ™). For each query, the search engine returns a result set of web pages. Each of the search result web pages are retrieved and these pages are themselves processed to download the web pages they reference. This "crawling" process continues recursively to a depth specified by the user. As pages are retrieved they are copied to the handheld computer, so they can be available offline. In this way, then, the user's searches are guaranteed to be performed, even if the user has forgotten about the exact notes they made, or has no time to do the searches manually.

Using the search results

After processing the queries, the results are made available to the user in a four ways that differ in terms of the device used to access the information, the need for an Internet connection, and the degree of interactivity between the user and the system. Due to the delay between sub

1.Offline on the handheld computer. They can view a list of all the queries recorded – along with the date and time each was input – and access the downloaded cached result sets and associated web pages (see Figure 3). Millions of handheld users are already using services such as Avantgo (www.avantgo.com) to read web pages offline; our approach extends this popular activity to search.

The pilot study of some aspects of the system design (Jones et al, 2003b) further illustrated the difficulties users face when accessing pages designed for the large screen, in the reduced space these devices provide. To help reduce the frustration users have in reading Web pages on the small screen, we have investigated the use of automatically extracted keyphrases. The aim is to provide users with concise, succinct descriptions of the pages retrieved, reducing the amount of time spent navigating the documents themselves (Jones et al, 2003c). We compared how well users are able to classify documents

when they are given just a list of keyphrases as a description of the content versus their performance when then have only the author written document titles (an example of the two presentations is shown in Figure 4). The results indicate that keyphrases perform as well as titles overall, while having the advantage of being able to be generated automatically: where documents have no, or poor author specified titles (often the case in the web) and in cases where the titles rely heavily on domain knowledge, the keyphrase approach appears to have some advantages. Furthermore, the keyphrases used in the search result lists (e.g. Figure 4) can be used when the full document is displayed to aid skim reading (see Figure 5).

2. Online on the handheld computer – if the user has a wireless connection available, they can access pages, linked to by the cached result pages, but not downloaded during the processing stage; and, perform further searches stimulated by the cached results.

3. Online on a desktop computer. We have a purpose-built browser for desktop searching that allows the users to view the handheld captured queries; access (online) the result sets produced by these queries; and, to carry out further searches online, thereby interactively refining the results of the initial queries.

4. Background, ambient use of search results on the desktop and large interactive displays (LIDS). We are also adapting the Collage Machine (Kerne, 2000) to display the results of queries gathered via the handheld over time on the desktop or our large interactive displays (Apperley & Masoodian, 2000). As Figure 3 illustrates, images are extracted from web pages relating to all of the search queries, and are collaged together[2]. A user can "click" on any image, and the relevant Web page is then displayed. The adaptations we are exploring, include: grouping images together based on the query they relate to (see Figure 4); enabling text as well as images to be effectively collaged; providing support for collaborative

[2] A demonstration of collaging can be found at http://www.cs.waikato.ac.nz/~mattj/jasisDemo.html

collages created by multiple users and their handheld devices; and, providing collage facilities for queries to Greenstone Digital Library collections (Witten et al, 2001). Again, the aim is to explore the benefits of a "calmer" approach to search that does not demand instant attention.

RELATED APPROACHES

Our interest is in supporting people in their background information seeking – both in terms of capturing queries and reviewing them, and the results, later. In contrast, others have proposed approaches where the system, rather than the user, works in the background. For example, Aridor *et al* (2001), have demonstrated an offline/online handheld search system that "pushes" information from a PC-based intelligent agent to a handheld device.

The XLibris system (Shilit et al, 1999) is a sophisticated electronic document reading support tool. As the user annotates the text, with a stylus, the system generates links to additional documents that might interest the reader. Phrasier (Jones & Staveley, 1999) provides similar facilities but automatically extracts keyphrases from each document and uses these, rather than annotated phrases, to suggest related material.

While systems like Xlibiris provide searches focussed around documents, we are interested in searches that arise within mobile situations. Other researchers have looked at linking physical contexts while mobile with later online information access. The InfoPoint (Kohtake et al, 2001), for instance, is a hand-held device that allows users to capture information from objects tagged with a visual code. This information can be used, later, for electronic content access when the user returns to their workstation.

Other qualities of are approach have parallels too. Our desire to support slower, more reflective approaches to search have been motivated by the concept of "slow technology" (e.g. Hallnas & Redstrom, 2001) .The potential benefits of ambient result display, related to our collaging approach, are

explored further in (Churchill et al, 2003). Finally, we use a handheld device and a personal computer together asynchronously; others, for instance, Myers and his Pebbles groups (Myers, 2001) have combined devices in synchronous ways.

Our work has been based on studying "knowledge workers". However, there are examples of the use of batched or "delayed" search result provision to support users in developing countries who do not have high-speed, reliable internet connections (Thies et al, 2002).

REFLECTIONS ON BEST PRACTICE

From this work, there are some general lessons for others who wish to develop search type applications using mobile devices:

- Developers should not simply view mobile handsets as standalone, "do everything" devices. Rather, an "ensemble" of devices can support the activity of searching, over an extended period of time.
- While there is a trend towards "always on" mobile devices, there is clearly a role in search for intermittent connectivity, too.
- Mobile devices present interesting challenges for search tasks. We saw, in particular, the importance of careful design of search result lists and the presentation of result pages.

CONCLUSIONS AND FUTURE WORK

Researchers have studied search as a "foreground" activity extensively. In contrast, there is far less work that explores the possibilities of supporting searches that occur in the "background". Our prototype illustrates how users might follow-up interests that arise while they are engaged in other activities. Clearly, there are many interesting open questions such as how the delay between "submitting" a query and receiving the results may impact on the way users' search questions evolve (Taylor 1968). We are

about to embark on a series of user studies of the tool to investigate such issues over an extended period and are also planning on evaluating the role such a tool might play in building collaborative communities of searchers.

ACKNOWLEGEMENTS

Thanks to Sebastian Dusterwald and Brendan Waugh for their earlier trial implementations of the prototype. David Bainbridge has assisted us with the collaging implementations.

REFERENCES

Apperley, M. and Masoodian, M. (2000): "Supporting collaboration and engagement using a whiteboard-like display", in Shared Environments to Support Face to Face Collaboration: A CSCW Workshop, Philadelphia, Pennsylvania, p 22-26

Aridor, Y., Carmel. D., Maarek, Y. S., Soffer, A. & Lempel, R. (2001). Knowledge encapsulation for focused search from pervasive devices. ACM Transactions on Information Systems, 20(1): 25-46.

Brown, B., Sellen, A. J. & O'Hara, K. P. (2000). A diary study of information capture in working life. In Proceedings of ACM CHI 2000, 438-445. ACM Press.

Churchill, E. F., Nelson, L. & Denoue, L. (2003). Multimedia Fliers: Information Sharing With Digital Community Bulletin Boards In Proceedings of Communities and Technologies 2003, September 2003, Kluwer Academic Publishers.

Hallnäs, L. & Redström, J. (2001). Slow Technology – Designing for Reflection. Personal and Ubiquitous Computing (5)3, 201-212.

Jones M., Buchanan, G. & Thimbleby, H. (2002). Sorting out searching on small screen devices. In Paterno, F. (Ed.), Proceedings of the 4th International Symposium on Mobile HCI, Pisa, Italy, September 2002, LNCS 2411, pp 81-94. Springer.

Jones, S., & Staveley, M. (1999). Phrasier: a System for Interactive Document Retrieval Using Keyphrases. In M. Hearst & F. Gey & R. Tong (Eds.), Proceedings of SIGIR'99: the 22nd International Conference on Research and Development in Information Retrieval (pp. 160-167). Berkeley, CA: ACM Press.

Jones, M., Buchanan, G. & Thimbleby, H. (2003a). Improving web search on small screen devices. Interacting with Computers, 14(2003) 479-495.

Jones, M., Jain, P., Buchanan, G. & Marsden, G. (2003b). Using a Mobile Device to Vary the Pace of Search. Proceedings of 4th International Symposium on Mobile HCI, Springer (forthcoming, September 2003).

Jones, S, Jones, M. & Deo, S. (2003c). Using Keyphrases as Search Result Surrogates on Small Screen Devices. CS Working Paper 07/03.

Kerne, A. (2000) "CollageMachine: an interactive agent of Web recombination." Leonardo, Vol. 33, No. 5, pp. 347–350.

Kidd, A., (1994). The Marks Are on the Knowledge Worker, Proc. CHI 94: Human Factors in Computing Systems, ACM Press, New York, 1994, pp. 186--191.

Kohtake, N., Rekimoto, J. & Yuichiro, A. (2001). Infopoint: a device that provides a uniform user interface to allow appliances to work together over a network. Personal & Ubiquitous Computing 5: 264-274.

Myers, B. (2001). Using Hand-Held Devices and PCs Together Communications of the ACM, 2001, 34-41.

Taylor, R. (1962). The process of asking questions. Journal of American Documentation, 13:391-396.

Taylor, R.S (1968). Question-negotiation and information seeking in libraries. College & Research Libraries, 29 (3). 178-194.

Schilit, B. N., Price, M. N, Golovchinsky, G., Tanaka, K. & Marshall, C. C (1999). As We May Read: The Reading Appliance Revolution. Computer, Vol. 32, No. 1, January 1999, pp. 65-73.

Thies, W., Prevost., J., Mahtab., T., Cuevas., G.T., Saad Shakhshir, Artola., A., Vo., B.D., Litvak., Y., Chan., S., Henderson., S. and Halsey, M., (2002). Searching the World Wide Web in Low-Connectivity Communities. Proceedings of 11th International World Wide Web Conference (available from: http://www2002.org/CDROM/alternate/714/index.html)

Thomas, P., Macredie, R. & Nardi, B. (2002). "The New Usability". Special Issue of ACM Transactions on Human Computer Interaction, June 2002.

Witten, I.H., Bainbridge, D., and Boddie, S.J. (2001). "Greenstone: open source digital library software". Communications of the ACM, vol 5, pp 47.

FIGURES

Figure 1: handheld application for query capture. User is entering a new query and a list of all previously entered queries is shown. Some of the queries (e.g. "site:www.microsoft.com digital ink") were entered using the "Advanced Search" interface.

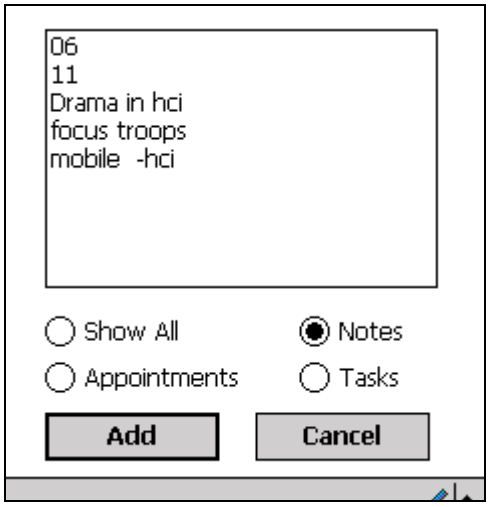

Figure 2: modified handheld query capture tool automatically gathers potential queries from users other handheld applications. Here the queries from "notes" are shown.

Figure 3: accessing offline search results. Selecting any search query (e.g. "digital photo frame") leads user to the search result list for that query; from there the user can view the related cached web pages.

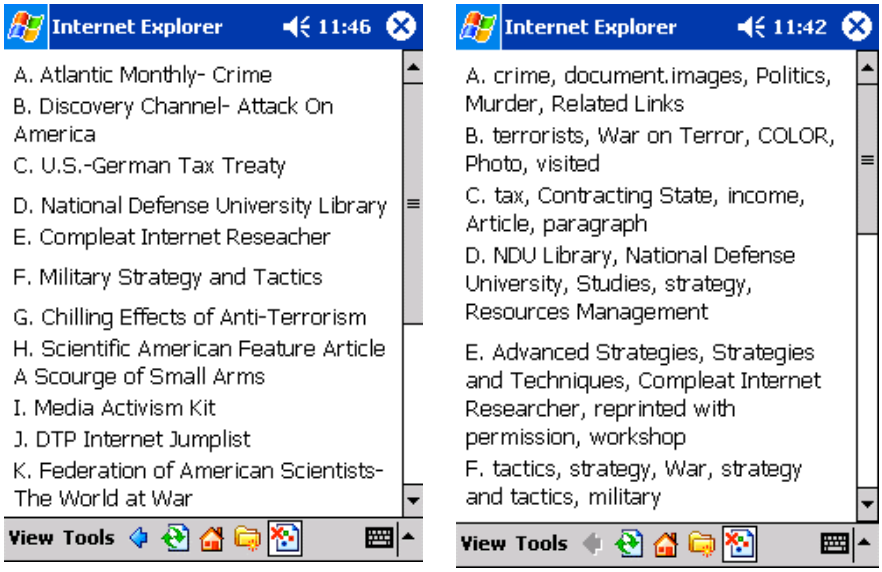

Figure 4: Example search result list presentations with: document titles only (left-hand-side) and automatically extracted keyphrases only (right-hand-side) on small screen of a Pocket PC (Jones et al, 2003c)

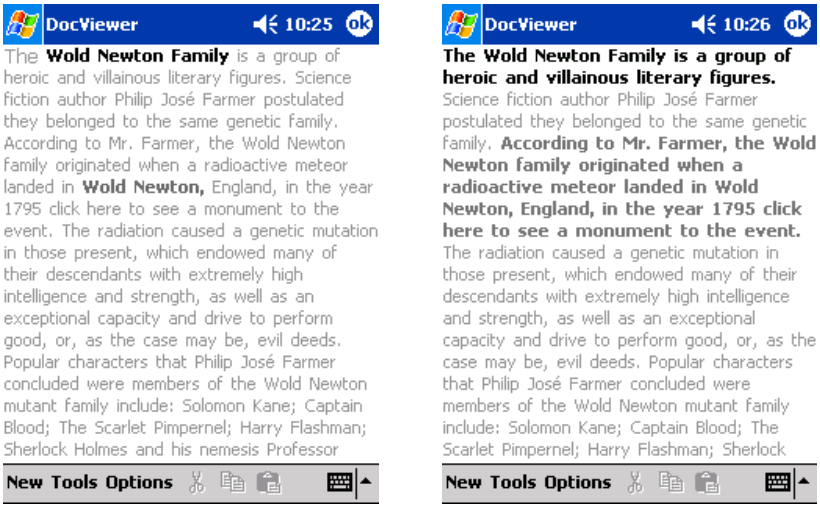

Figure 5: prototype skimming interface (Jones et al, 2003c), showing keyphrase emphasis (left) and sentence emphasis (right)

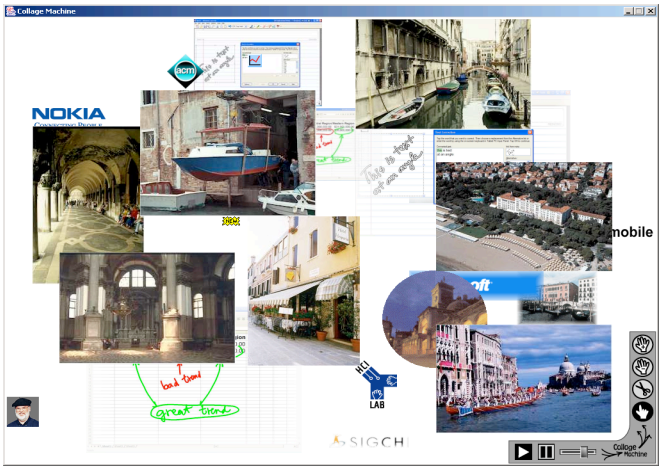

Figure 6: collage created by the Collage Machine (Kerne, 2000) given search queries captured by the handheld device.

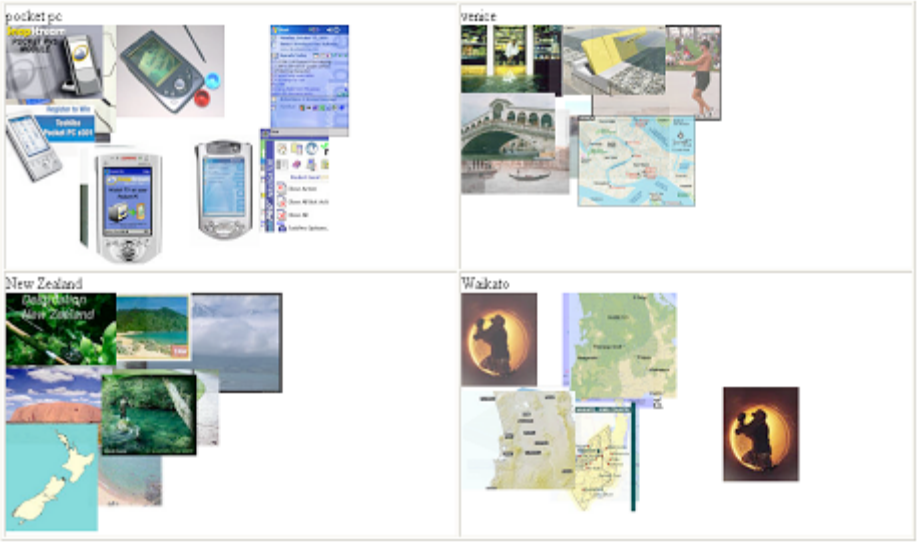

Figure 7: collage created by our system (and displayed on a desktop sized screen). Four regions are shown, in each, images relating to one search query (e.g. “New Zealand”) are being collaged together. Overtime, the display dynamically changes to show images for all of the users, “background” queries.